STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–10.
Published online in International Academic Press (www.IAPress.org)

A Comparative Statistical Analysis of Decision Tree and AdaBoost Ensemble for Employee Performance Classification in the Hospitality Sector

Teguh Herlambang[1,2], Mohamad Yusak Anshori[3,4], Bambang Suharto[5*], Zuraini Othman[6], Mohd Sanusi Azmi[7], Mochammad Romli Arief[8]

[1] *Department of Information System, Faculty of Economics Business and Digital Technology, Universitas Nahdlatul Ulama Surabaya, Indonesia*
[2] *Center for Data and Business Intelligence, Universitas Nahdlatul Ulama Surabaya, Indonesia*
[3] *Departement of Management, Universitas Nahdlatul Ulama Surabaya, Surabaya, Indonesia*
[4] *Research Group of Intelligence System and Financial Technology, Universitas Nahdlatul Ulama Surabaya, Indonesia*
[5] *Department of Tourism and Hospitality, Faculty of Vocational, Airlangga University, Surabaya, Indonesia*
[6] *Department of Diploma Studies, Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka, Malaysia*
[7] *Department of Software Engineering, Universiti Teknikal Malaysia Melaka, Malaysia*
[8] *Undergraduate Department of Information Systems, Universitas Nahdlatul Ulama Surabaya, Surabaya, Indonesia*

Abstract This study aims to classify the factors affecting employee job satisfaction in Hotel X using the Decision Tree (DT) and Adaptive Boosting (AdaBoost) methods. The hospitality industry relies heavily on human capital to deliver high-quality services, and employee satisfaction is directly linked to service excellence, loyalty, and organizational performance. Data were collected from Hotel X's internal Employee Satisfaction Index (ESI), comprising 70 records and 9 response indicators across multiple departments. Exploratory Data Analysis (EDA), correlation analysis, and label encoding were performed to prepare the dataset. The Decision Tree was first utilized to model the classification of employee satisfaction levels, followed by optimization using the AdaBoost ensemble method to enhance predictive accuracy. Three simulations were conducted using training-to-testing ratios of 70:30, 75:25, and 80:20, respectively. The results show that AdaBoost consistently improved the classification performance, achieving the highest accuracy of 93% in the third simulation. These findings underscore the significance of ensemble learning techniques in enhancing model reliability for human resource analytics in the hotel industry. This research demonstrates the practical value of combining DT and AdaBoost for workforce data analysis in service-based organizations. The model can be adapted to various industries that prioritize employee satisfaction as a key performance driver.

Keywords Hotel employee satisfaction; Decision Tree; AdaBoost; Classification; Ensemble learning; Hospitality analytics

DOI: 10.19139/soic-2310-5070-3493

1. Introduction

Hotels are the world's leading employers in the tourism industry, and the high degree of mobility or turnover in this industry is a global concern [1]. This situation is certainly not convenient for hotel management, as it can disrupt services. As is well known, hotels are service-based companies. Service quality impacts guests' satisfaction through providing service and performance [2]. Therefore, with excellent coordination and synergy among employees, hotels have the opportunity to generate significant organizational benefits and foster a high sense of employee loyalty.

Work motivation plays a critical role in shaping employee behavior and can be a pivotal factor in fostering employee loyalty. Employees with high work motivation tend to be more enthusiastic in performing their duties, exhibiting better mental well-being, and show a stronger desire to work optimally [3]. Loyalty is an emotional

*Correspondence to: Bambang Suharto (Email: bambang.suharto@vokasi.unair.ac.id).
Department of Tourism and Hospitality, Faculty of Vocational, Airlangga University, Surabaya, Indonesia.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

construct [4] that can be shaped by leadership behavior and organizational culture. An open and communicative leadership style helps build strong relationships between leaders and subordinates. When employees are involved in decision-making processes and feel trusted, their commitment to the organization tends to increase [5].

Employees are widely regarded as the most essential asset in the hospitality industry [6]. Human resources serve as the driving force behind achieving a hotel's strategic goals. The growth and success of the hospitality sector is closely linked to workforce performance, which underscores the importance of human capital development [7]. Unlike in manufacturing industries—where machines handle most tasks, the hotel industry relies heavily on human interaction for tasks ranging from guest reception to final farewells [8].

Findings from prior research confirm the strong, positive correlation between job satisfaction, employee loyalty, and job performance [9]. In the context of sustainable corporate growth, job satisfaction is particularly vital to reducing employee turnover and enhancing organizational loyalty. Moreover, job satisfaction has a well-documented impact on product quality, service delivery, customer relationships, and organizational performance [10]. It has been shown to be a major driver of productivity, efficiency, and profitability across business settings [11]. Job satisfaction should no longer be seen merely as an internal HR matter, but rather as a strategic imperative linked directly to organizational competitiveness. In the hospitality sector, understanding the dynamics of employee satisfaction is essential not only for employee retention but also for maintaining service quality and brand reputation.

The development of information technology and big data analytics now allows organizations to scientifically identify and forecast the factors influencing employee satisfaction in the hospitality sector. Data-driven decision-making, business intelligence, and analytics (BI&A) have been shown in many studies to provide a competitive edge for organizations [12]. Business intelligence functions as a technology for transforming raw data into actionable insights, encompassing data mining, visualization, analytical tools, and infrastructure to enhance decision-making quality and organizational profitability [13].

With the help of artificial intelligence, today's data processing capabilities are more sophisticated than ever. Machine learning, a subset of AI and computer science, enables systems to learn patterns from data and improve over time without being explicitly programmed [14]. These algorithms can be deployed for tasks such as prediction, classification, and clustering. Classification algorithms are trained on labeled datasets to recognize patterns and make predictions on new data [15].

This study focuses on classifying the factors that influence job satisfaction among employees at Hotel X using the Decision Tree and AdaBoost method based on primary data. As additional information, this study is a continuity from previous hotel case studies related to occupancy [16] and revenue [17],[18]. Decision Trees are one of the most widely used methods in classification problems. They work by splitting the dataset from root to leaf nodes based on decision rules, eventually leading to a class label [19]. As a hierarchical, non-parametric model, Decision Trees are highly effective under conditions of uncertainty and can partition data into homogeneous subsets for classification purposes [20]. AdaBoost, short for Adaptive Boosting, is a widely recognized ensemble method used to improve the performance of classification models, particularly on imbalanced datasets [21]. It has led to extensive theoretical exploration in both machine learning and statistical domains [22].

Decision Trees have been successfully implemented in various domains. For instance, in 2020, Apriliani et al. used the Decision Tree to analyze hotel service sentiment, achieving an accuracy of 88.54% [**?**]. In 2022, Shrestha et al. used it to analyze and predict tourist satisfaction, with a precision score of 0.879 [**?**]. Moreover, previous research has demonstrated the applicability of the Decision Tree method in diverse areas such as healthcare, banking, and beyond.

2. Research Methods

2.1. Data Acquisition

The dataset used in this research is taken from the historical private data of Hotel X's ESI assessment. In conducting employee assessments, hotel management utilized questionnaire media and collected a dataset limited with composition of 70 rows and 9 columns. Data is separated based on the accumulation of responses from 10 categories that have been done by employees from each department. Below in Table 1 is dataset that have been processed and Figure 1 is flowchart of research method.

Table 1. Dataset

Name of Department	Type of Question	Always	Frequently	Sometimes	Rare	Never	Result	Category
HR	Department	1	1	9	1	0	3.17	Low
	Pride	4	2	1	1	1	3.78	Low
	Job	16	5	1	2	0	4.46	Standard
	Cooperation	13	5	3	0	0	4.48	Standard
	Care	13	2	2	1	0	4.50	Standard
	Leadership	23	4	2	1	0	4.63	Standard
	Communication	9	4	2	0	0	4.47	Standard
	Confession	9	1	1	1	0	4.50	Standard
	Benefit	9	4	0	0	0	4.60	Standard
	Personnel	19	5	0	0	0	4.79	Standard
			...					
Food and Beverages	Department	9	10	1	0	0	4	Standard
	Pride	8	4	3	0	0	3	Low
	Job	14	21	2	2	1	8	High
	Cooperation	23	8	4	0	0	7	High
	Care	18	6	6	0	0	6	Standard
	Leadership	28	19	3	0	0	10	High
	Communication	17	7	1	0	0	5	Standard
	Confession	14	1	5	0	0	4	Standard
	Benefit	10	7	8	0	0	5	Standard
	Personnel	17	16	2	4	1	8	High

Based on Table 1, it can be observed that the assessment was conducted on seven departments with ten categories of questions related to work and the company. The question categories contained in the questionnaire have five answer variations that adjusted to each individual's opinion. The assessment of each individual, then accumulated and divided to find the result in the form of an average (mean) value. From the results, the average value is then categorized into three criteria, namely low, moderate, and high. Below in Figure 1 is a flowchart of the research method utilized to carry out this study.

Based on Figure 1, it can be observed that the classification process in this study was carried out through several stages. Of the stages mentioned above, the preprocessing analysis stages are crucial. At this stage, the data will be processed to identify its relationship, patterns, distribution, and other aspects, thus determining the appropriate classification model and technique to be applied. In the final stage of Figure 1 above, it can also be observed that the classification results simulated by the model will be evaluated to find the level of accuracy based on a certain method and displayed through a diagram.

2.2. Exploratory Data Analysis (EDA)

EDA is a technique used to explore datasets so as to extract useful and actionable information, identify relationship among the explanatory variables, detect mistakes, and preliminary select appropriate models [23]. From the EDA process that has been carried out, some information is obtained as shown in Figures 2–4 below.

Based on Figure 2, it can be observed that the moderate category is the most numerous or dominant. This indicates that the environmental conditions, and other supporting aspects are fairly good. Therefore, from this diagram, it can tentatively conclude that employees at this hotel are quite satisfied with their current employer. Next, from Figure 3, it also can be observed that department of HR has dominant moderate responses, followed by engineering department. Furthermore, the housekeeping department achieved the most low responses. This requires special attention from the company to improve the performance of these departments.

From Figure 4 above, it can be observed that the leadership aspect received the best response and be the highest, followed by responses related to personnel. This indicates that both aspects were rated the highest by all departments and employees. This demonstrates that these two aspects are key to employee satisfaction at work.

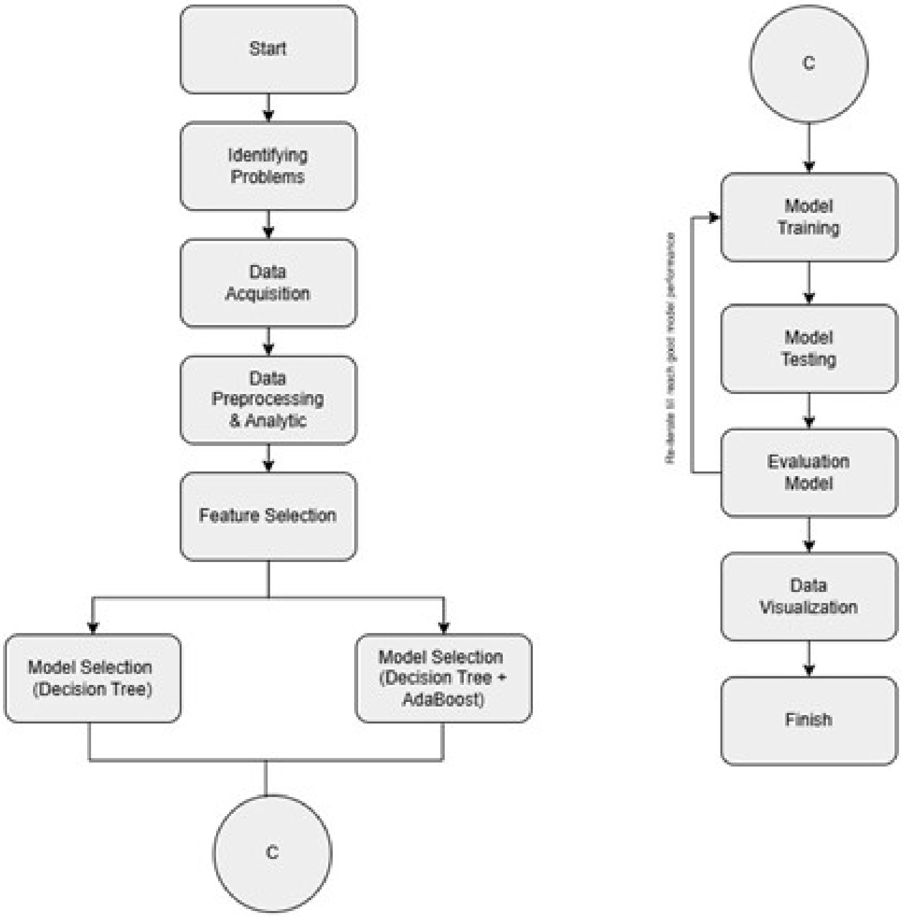

Figure 1. Research Methodology

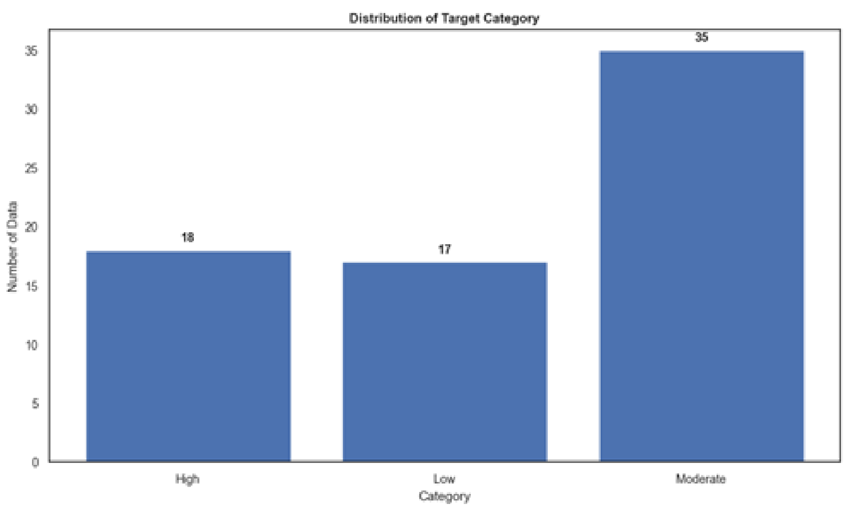

Figure 2. Distribution of Target Category

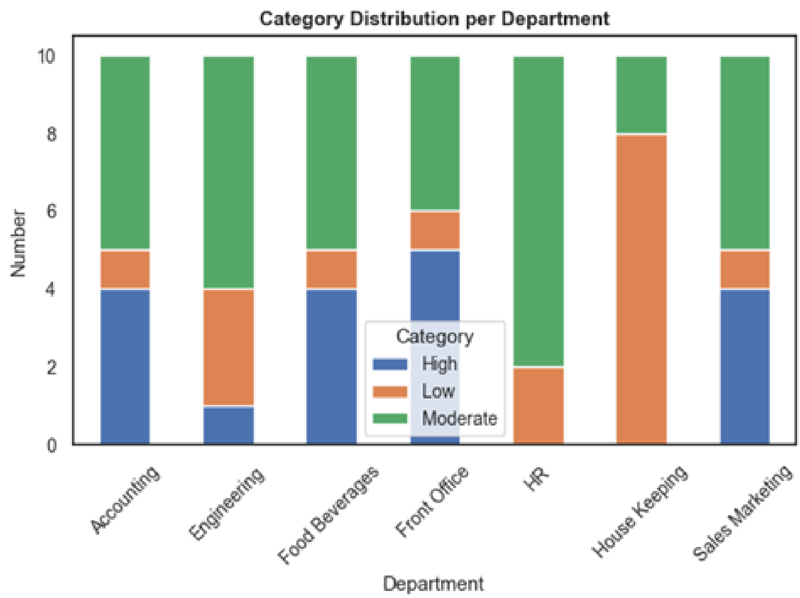

Figure 3. Distribution Category per Department

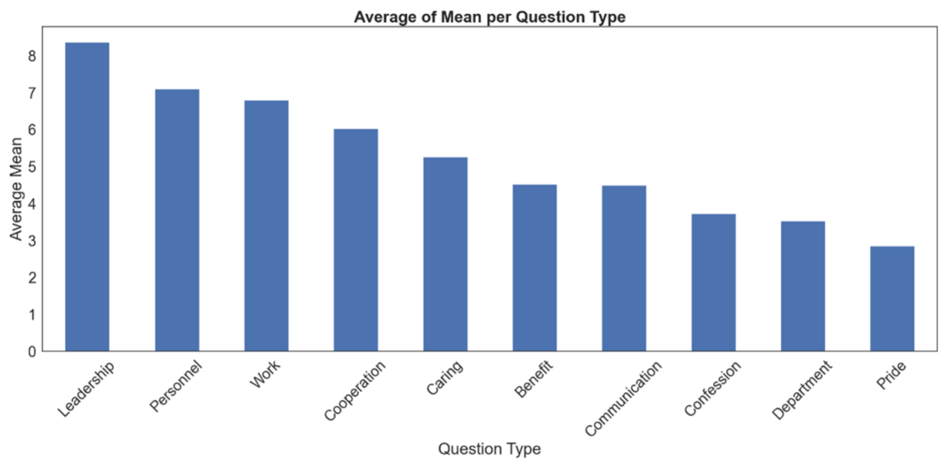

Figure 4. Distribution of Mean per Question Type

2.3. *Statistical Analysis*

For statistical analysis is started by implementation of correlation analysis. In the correlation analysis stage, the independent variables are measured for their impact on the dependent variable using Spearman's rule. Spearman's correlation coefficient determines a simple linear relationship between two variables and measures without dimensions [24]. Below in Equation (1) is function of Spearman's rule and Figure 5 is the result of correlation.

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{1}$$

where:

- n: Number of data pair

- $\sum d_i^2$: Sum of squares of the rank difference

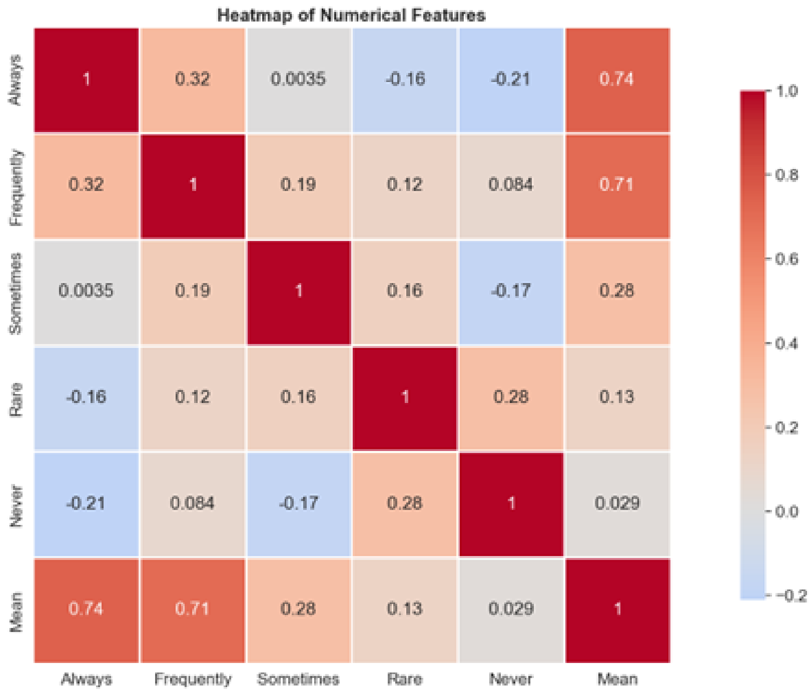

Figure 5. Correlation Diagram

From Figure 5, it can be observed that feature always and frequently where both features have high correlation value to the mean feature than the other features and indicated be the most importance features. Furthermore, below in Table 2 are the result of other statistical tests

Table 2. Result of Encoding Label

Type of Statistical Test	Result
Normality (Shapiro-Wilk)	Distribution of Mean feature is not normal with p-value of 0.0094.
Difference Between Feature (Kruskal-Wallis)	There is significant difference between Mean and Category features with p-value of 0.00
Association Between Department and Category Features (Chi-Square)	There is significant association between Department and Category features with p-value of 0.0019.

2.4. *Encoding Label*

In most case studies related to classification, the encoding process is crucial. It has to be applied because the data utilized and processed are in the text form, which requires transformation. Label encoding is the process of converting the label of text/categorical values into a numerical format that ML algorithms can interpret [25]. In this case, all values in target column convert to numeric format as follows in Table 3.

2.5. *Feature Selection*

Below in Table 4 containing variable independent and dependent that determined to be utilized. The selection of features in Table 4 below was based on the primary responses provided by employees in the questionnaire. Other features, such as name of department and type of question were eliminated due to their repetitive nature.

Table 3. Result of Encoding Label

Value of Target Variable	Converted Result
High	0
Low	1
Moderate	2

Table 4. Feature Selection

Variable Independent	Variable Dependent
Name of Department	Category
Always	
Frequently	
Sometimes	
Rare	
Never	

2.6. *Decision Tree Classification Method*

The Decision Tree algorithm is a method used to learn and predict patterns within data and represent the relationship between attribute variables in the form of tree structures [26]. Below in Equations (2)–(3) is function of Decision Tree algorithm.

$$Entropy(S) = \sum_{i=1}^{m} -p(w_i \mid S) \log_2 p(w_i \mid S) \tag{2}$$

Notes:

- S: Set of cases being analyzed
- m: Total number of different classes within the data set S
- $p(w_i \mid S)$: Probability of occurrence of class w_i in dataset S

After that, the next step is to calculate the gain value, a measure of how much information is obtained from separating data based on certain attributes. The gain calculation function is as follows:

$$Gain(S, J) = Entropy(S) - \sum_{i=j}^{n} p(v_i \mid S)\, Entropy(S_i) \tag{3}$$

Notes:

- S: Set of cases being analyzed
- J: Features/attributes considered in data separation
- n: Number of classes in the node
- $p(v_i \mid S)$: Proportion of v values appearing in the class in the node
- S_i: Entropy of the composition of v values for the j-th class in the i-th data node

2.7. *AdaBoost Model*

Adaptive Boost is one of the supervised machine learning techniques that relates to a spesific method to learn a booster classifier. It is classification method to construct strong learners from a linear combination of weak learners [28]. Below in Equation (4) is function of AdaBoost method.

$$f_t(x) = \sum_{i=1}^{C} a_t h_t(x) \tag{4}$$

Where f_t is a weak linear relationship and $a_t h_t$ as the set of weak learners that are considered the last classifier.

2.8. *Evaluation Metrics*

The assessment method is a key factor in evaluating the classification performance and guiding the classifier modeling. There are three main phases of the classification process, namely, training phase, validation phase, and testing phase [28]. There are three metrics used to evaluate classification model, namely precision score, f1-score, recall score, and accuracy score. Below in Equation (5) is precision function, Equation (6) is f1-score function, Equation (7) is re-call function, and Equation (8) is accuracy score.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{f1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \tag{8}$$

3. Results and Discussion

At this stage, several steps are explained as preparation for carrying out classification simulations using Decision Tree and AdaBoost methods. As previously known, the dataset utilized in this study was limited, so some adjustments are required, including utilize Synthetic Minority Over-sampling Technique (SMOTE) and implementation of cross-validation, where in this study using 5-fold stratified. Below in Table 5 are the details of the results of the SMOTE application for imbalance dataset.

Table 5. Imbalance Check

Name of Class	Before SMOTE	After SMOTE
High (0)	18 samples	35 samples
Low (1)	17 samples	35 samples
Moderate (2)	35 samples	35 samples

Table 5 above illustrates the differences before and after the SMOTE technique was applied. The number of samples in the data increased, preventing the model from overfitting during the learning phase. Next, below in the Table 6 and Table 7 below are details regarding the hyperparameter tuning of the Decision Tree and AdaBoost methods. In Table 8 below also including the number of training and testing dataset after SMOTE technique applied.

Table 6. Hyperparameter Tuning of Decision Tree

Name of Hyperparameter	Value
max_depth	5
min_samples_split	10
min_sample_leaf	5
random_state	42
class_weight	balanced

Table 7. Hyperparameter Tuning of AdaBoost

Name of Hyperparameter	Value
estimator	Decision Tree
n_estimator	50
learning_rate	0.8
random_state	42

Table 8. Data Splitting After SMOTE

Percentage of Splitting	Number of Training Data	Number of Testing Data
70 : 30	73	32

From the adjustments explained in Table 6 to Table 8, the simulation results are obtained which can be observed in Figure 6, namely the confusion matrix and Figure 6, namely the model performance.

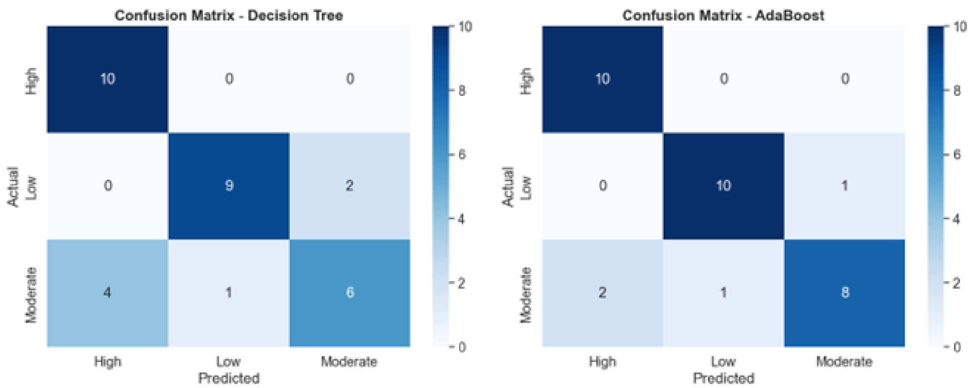

Figure 6. Confusion Matrix

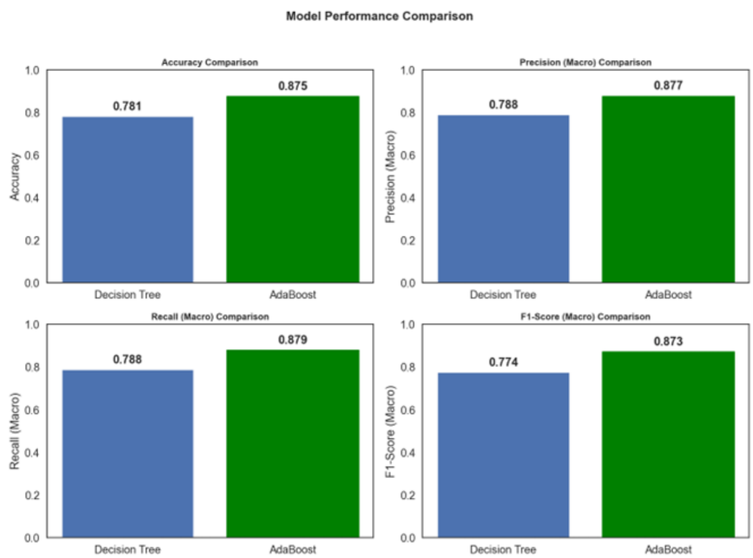

Figure 7. Model Performance

Figure 6 above illustrates the results of the classification simulation utilizing both methods. The classification results appear quite good, as indicated from the number of True Positive generated. The Figure 7 above presents a detailed breakdown of the metrics generated by both methods, illustrating the overall classification metric evaluation. For accuracy, the Decision Tree method achieved a score of 0.781, or approximately 78%. The AdaBoost method achieved an accuracy score of 0.877, or approximately 88% and followed by other metrics such as Precision, Recall, and F1-Score, respectively. Furthermore, a significant comparison of the performance of the two methods is further demonstrated using Paired T-test and Wilcoxon signed-rank test in Table 9 below.

Table 9. Significant Difference

Result of Paired-T-test	Result of Wilcoxon signed-rank test
p-value: 0.6989	p-value: 0.6250

Table 9 above illustrates the results of the significant difference of both methods while conducting simulation and resulting classification. The result of comparison obtained by both method indicating that no significant difference occurred.

4. Conclusion

From the simulation results that have been carried out, overall it produces a good level of classification accuracy and also has fulfilled the objective of this study while conducted in the limited dataset. The accuracy value obtained by Decision Tree is approximately 78%, while AdaBoost reached 88 percent. For the next study, the authors will improvise with a larger number of datasets and more advanced methods.

5. Acknowledgment

The author would like to thank LPPM-Universitas Nahdlatul Ulama Surabaya (UNUSA) for their support in providing the facility for this research. Additionally, gratitude is extended to the Centre of Research and Innovation Management of Universiti Teknikal Malaysia Melaka (UTeM) for sponsoring the publication fees under the Tabung Penerbitan CRIM UTeM.

REFERENCES

1. N. D. Djajasinga, L. Sulastri, A. Sudirman, A. L. Sari, and E. L. Rihardi, "Practices in human resources and employee turnover in the hospitality industry," *Advances in Social Science, Education and Humanities Research/Advances in Social Science, Education and Humanities Research*, Jan. 2021, doi: 10.2991/assehr.k.210615.023.
2. B. J. Ali et al., "Hotel Service quality: The impact of service quality on customer satisfaction in hospitality," *International Journal of Engineering Business and Management*, vol. 5, no. 3, pp. 14–28, Jan. 2021, doi: 10.22161/ijebm.5.3.2.
3. U. Nuryanto, C. Warman, Suharno, E. Febrianata, R. Purnamasari, and D. E. Subroto, "The Roles of Employee Loyalty between Work Motivation, Job Satisfaction, Leadership Against Employee Performance," *IOSR Journal of Business and Management (IOSR-JBM)*, vol. 23, no. 2, pp. 15–25, 2021.
4. N. Z. Ena, A. E. Sjioen, and A. M. Riwudjami, "The Effect of Work Environment on Employee Loyalty with Work Stress as an Intervening Variable at Bella Vita Hotel - Kota Kupang," *Quantitative Economics and Management Studies*, vol. 3, no. 1, pp. 65–76, Jun. 2022, doi: 10.35877/454ri.qems865.
5. M. T. Andrianto, A. Moeins, and W. Sunaryo, "The influence of leadership style on employee performance and loyalty mediated by organizational climate," *Journal of World Science*, vol. 2, no. 8, pp. 1180–1187, Aug. 2023, doi: 10.58344/jws.v2i8.386.
6. K. M. Faraj, D. K. Faeq, D. F. Abdulla, B. J. Ali, and Z. M. Sadq, "Total quality management and hotel employee creative performance: the mediation role of job embeddedment," *SSRN Electronic Journal*, May 2021.
7. Y. Rivaldo, "Leadership and Motivation to Performance through Job Satisfaction of Hotel Employees at D'Merlion Batam," *The Winners*, vol. 22, no. 1, Mar. 2021, doi: 10.21512/tw.v22i1.7039.
8. J. Norbu and P. Wetprasit, "The study of Job Motivational Factors and its Influence on job Satisfaction for hotel employees of Thimphu, Bhutan," *Journal of Quality Assurance in Hospitality & Tourism*, vol. 22, no. 2, pp. 245–266, May 2020, doi: 10.1080/1528008x.2020.1769524.

9. T. T. K. Phuong and T. T. Vinh, “Job satisfaction, employee loyalty and job performance in the hospitality industry: a moderated model,” *Asian Economic and Financial Review*, vol. 10, no. 6, pp. 698–713, Jan. 2020, doi: 10.18488/journal.aefr.2020.106.698.713.
10. P. Heimerl, M. Haid, L. Benedikt, and U. Scholl-Grissemann, “Factors influencing job satisfaction in hospitality industry,” *SAGE Open*, vol. 10, no. 4, Oct. 2020, doi: 10.1177/2158244020982998.
11. M. Farrukh, “Role of job satisfaction and organizational commitment in employee loyalty: Empirical analysis from Saudi hotel industry,” *International Journal of Business and Psychology*, vol. 2, no. 1, 2020.
12. S. Siripipatthanakul, B. Phayaphrom, and P. Tong-On, “The implementation of business intelligence using data analytics and its effects towards on performance in the hotel industry in Thailand,” *SSRN Electronic Journal*, Oct. 2021.
13. E. D. Madyatmadja, C. N. A. Adiba, D. J. M. Sembiring, D. Pristinella, and A. M. Putra, “The positive impact of implementation business intelligence and big data in hospitality and tourism sector,” *International Journal of Emerging Technology and Advanced Engineering*, vol. 11, no. 6, pp. 59–71, Jun. 2021, doi: 10.46338/ijetae0621_07.
14. D. Mamadou, I. Sanou, and A. Sere, “An intelligent method of detecting and classifying hotels for an efficient choice,” *Revue Africaine De Recherche En Informatique Et Mathématiques Appliquées*, 2024.
15. A. Herrera, Á. Arroyo, A. Jiménez, and Á. Herrero, “Forecasting hotel cancellations through machine learning,” *Expert Systems*, vol. 41, no. 9, Apr. 2024, doi: 10.1111/exsy.13608.
16. M. Y. Anshori, T. Herlambang, and M. F. A. Yaziz, “Optimizing Occupancy of Hospitality Sector Using Support Vector Regression and Genetic Algorithm,” *Journal of Revenue and Pricing Management*, Jun. 2025, doi: 10.1057/s41272-025-00539-4.
17. M. Y. Anshori, P. Katias, T. Herlambang, N. S. Meutia, Z. B. Othman, and M. S. Azmi, “Predicting Hotel Revenue Using Gradient Boosting Regression and Support Vector Regression: A Comparative Analysis,” *Journal of Revenue and Pricing Management*, Sep. 2025, doi: 10.1057/s41272-025-00557-2.
18. M. R. Arief et al., “Forecasting revenue rate of hotel V by gradient boost regression and K-nearest neighbors (KNN) methods,” *AIP Conference Proceedings*, vol. 3372, p. 040019, Jan. 2025, doi: 10.1063/5.0299502.
19. A. Živković, D. Šebalj, and J. Franjković, “Prediction of the employee turnover intention using decision trees,” in *Proceedings of the 26th International Conference on Enterprise Information Systems*, 2024.
20. S. Y. Kim and A. Upneja, “Majority voting ensemble with a decision trees for business failure prediction during economic downturns,” *Journal of Innovation & Knowledge*, vol. 6, no. 2, pp. 112–123, Feb. 2021, doi: 10.1016/j.jik.2021.01.001.
21. Md. A. Jafor, Md. A. H. Wadud, K. Nur, and M. M. Rahman, “Employee Promotion prediction using improved AdABoost Machine Learning approach,” *AIUB Journal of Science and Engineering (AJSE)*, vol. 22, no. 3, pp. 258–266, Dec. 2023, doi: 10.53799/ajse.v22i3.781.
22. T. Islam et al., “Review analysis of Ride-Sharing applications using Machine learning approaches,” in *CRC Press eBooks*, 2023, pp. 99–122, doi: 10.1201/9781003253051-7.
23. S. Aldera, A. Emam, M. Al-Qurishi, M. Alrubaian, and A. Alothaim, “Exploratory data analysis and classification of a new Arabic online extremism dataset,” *IEEE Access*, vol. 9, pp. 161613–161626, Jan. 2021, doi: 10.1109/access.2021.3132651.
24. K. Ali Abd Al-Hameed, “Spearman’s correlation coefficient in statistical analysis,” *International Journal of Nonlinear Analysis and Applications*, vol. 13, no. 1, pp. 3249–3255, 2022, doi: 10.22075/ijnaa.2022.6079.
25. N. Ahmed et al., “Machine learning based diabetes prediction and development of smart web application,” *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 229–241, Jun. 2021, doi: 10.1016/j.ijcce.2021.12.001.
26. D. S. Suprapto and R. Oetama, “Analysis of airline passenger satisfaction using Decision Tree and Naïve Bayes algorithms,” *Jurnal Informatika Ekonomi Bisnis*, pp. 1493–1500, Dec. 2023, doi: 10.37034/infeb.v5i4.728.
27. S. N. Alsubari, “Data Analytics for the identification of fake reviews using supervised learning,” *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, vol. 70, no. 2, pp. 3189–3204, Sep. 2021, doi: 10.32604/cmc.2022.019625.
28. A. Tharwat, “Classification assessment methods,” *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, Aug. 2018, doi: 10.1016/j.aci.2018.08.003.