



Analysis of Study Space Comfort Levels in Coffee Shops Using Decision Tree and Random Forest Algorithms

Fitriany Aisyah Nurvella¹, Putri Ayu Dita Widya², and Fitri Nur Wahyuni³

⁴Michel Archie Satria Aji Prasetyo, ⁵Derviola Putri Cahayani, ⁶Az Zahra Ailsa Rasyid

¹²³Data Science of Universitas Pembangunan Nasional “Veteran” Jawa Timur,

⁴Nursing of Istanbul Kent University, ⁵Social Welfare of Jember University, ⁶Sociology Study of Airlangga University.

¹25083010049@student.upnjatim.ac.id, ²25083010070@student.upnjatim.ac.id,
³25083010076@student.upnjatim.ac.id ⁴2502090080@ogr.kent.edu.tr ⁵rvila10p@gmail.com
⁶azzahraailsar@gmail.com

ABSTRACT

Coffee shops are no longer just places to enjoy drinks, they have also become workspaces and study spaces for university students. This study examines this shift through Exploratory Data Analysis (EDA) of questionnaire data from 150 valid respondents, covering 22 variables such as visit frequency and duration, spending, number of devices, seating area, preferred time, noise disturbance, Wi-Fi and power outlet problems, and the percentage of tasks completed. The data were cleaned and checked for missing values and outliers using the IQR method. The results show that most respondents feel comfortable, with scores of 4 and 5 being the most common, particularly among students from Computer Science and Engineering and Science, in the late afternoon, and in workspace areas. Comfort is positively related to productivity, with the percentage of tasks completed rising from 52.0% for Uncomfortable respondents to 64.9% for Comfortable respondents and 74.1% for Very Comfortable respondents. However, Wi-Fi and power outlet problems remain common, experienced occasionally by 41.3% and often by 33.3% of respondents. Decision Tree and Random Forest models reached 50.0% and 56.7% accuracy respectively in predicting comfort, so they are not yet reliable without further tuning. Overall, the findings confirm that coffee shops now serve as student workspaces, and facility quality is a factor owners should address.

Keywords: Coffee Shop, Decision Tree, Random Forest, Comfort Levels

I. INTRODUCTION

In recent years, the function of coffee shops has undergone a significant shift, evolving from traditional social spaces into modern alternative workspaces for students and professionals. Following the Coronavirus Disease 2019 (COVID-19) pandemic, many cafés adopted a compound-space operation model, offering diverse consumption scenarios that cater to both leisure and work [1]. Driven by the explosive growth of remote work, more people now rely on these spaces for studying and employment [1]. Consequently, customer demographics often skew toward younger demographics, for instance, undergraduate students frequently make up the majority of users, with participation decreasing as education levels rise [2]. Ultimately, coffee shops are no longer evaluated solely on product quality or pricing, but increasingly on their capacity to function as comfortable work areas equipped with adequate facilities, stable internet, and conducive ambient conditions [3].

* Corresponding author.

E-mail address: 25083010076@student.upnjatim.ac.id

<http://doi.org/10.33005/jasid.v2i2.83>

However, transforming a commercial beverage venue into a functional study space introduces various complexities. Factors such as environmental noise, seating arrangements (indoor versus outdoor spaces), facility constraints, and visit duration significantly affect a visitor's overall comfort and task completion, as shown in studies of student behavior and seating choices in café settings [4]. The physical conditions of the café such as the seating arrangement show neutrality as well as offering flexible spaces for different kinds of interactions [5]. In an interview study of Indonesian university students, atmosphere, facilities such as Wi-Fi, and coffee were the factors that most influenced their choice to complete assignments in coffee shops, with ambience mentioned most often [6]. Identifying the spatial elements that foster user comfort in these establishments is essential for both commercial management and environmental design, as students exhibit varying levels of sensitivity to acoustic disturbances and furniture configurations when selecting their study environments [7]. While conventional reviews offer subjective feedback, machine learning classification models enable a quantitative, data-driven examination of the specific environmental factors that dictate consumer choices and comfort in alternative workspaces.

Machine learning techniques, particularly classification algorithms, offer robust frameworks for analyzing complex survey and operational data regarding consumer satisfaction and preferences. Among these, Decision Tree and Random Forest algorithms are frequently implemented for predictive modeling and feature importance analysis [8]. The Decision Trees remain vital for evaluating user satisfaction due to their transparent model structures, which provide managers and policymakers with clear, logical insights necessary for accountable decision-making within educational settings [8]. Meanwhile, Random Forest serves as a powerful ensemble counterpart to evaluate complex non-linear relationships and ensure robust predictive performance [10].

This study aims to analyze and predict the comfort levels of individuals utilizing coffee shops as study spaces by implementing Decision Tree and Random Forest algorithms. Specifically, this research places a primary focus on the Random Forest model to ensure robust predictive performance and meticulously identify critical attributes, such as environmental preferences and facility constraints that dictate user comfort, through feature importance analysis. The findings are expected to contribute comprehensive empirical insights into the modern shift of coffee shop utilities and offer actionable recommendations for workspace optimization.

Previous studies on cafés as study environments have largely relied on interviews and perception-based surveys [4]-[7], whereas Decision Tree and Random Forest have mainly been applied to satisfaction in academic information systems and public services [8]-[10]. To the best of our knowledge, few studies have used these algorithms to classify comfort levels in coffee shops used as study spaces. This study addresses that gap using survey data from Indonesian university students, covering visit behavior (duration, spending), facility conditions (Wi-Fi, power outlets), and noise disturbance. In addition, comfort levels are examined against the percentage of tasks completed, so that the findings can inform facility improvements by coffee shop owners.

II. METHODOLOGY

This section outlines the systematic workflow, experimental design, and analytical techniques employed to conduct the predictive analysis of coffee shop workspace comfort. To ensure a structured and replicable investigation, the research process is divided into six main stages: data collection through Exploratory Data Analysis (EDA) of survey responses, data preprocessing, machine learning model implementation, model training, performance evaluation metrics, comprehensive result analysis, and flow chart, as detailed in the following subsections.

A. Data Collection

The data collection phase details the primary dataset utilized in this study, which was gathered through an online survey to capture behavioral and environmental preferences regarding coffee shop workspaces.

TABLE I. SURVEY ATTRIBUTES AND DESCRIPTION

No.	Attribute	Data Type	Description
1.	In a week, how many times do you visit a coffee shop to do assignments or work?	Numeric (Discrete)	Visit frequency per week
2.	How comfortable are you when doing tasks at a coffee shop?	Categorical	Target variable (Comfort level)

No.	Attribute	Data Type	Description
3.	In a single visit, on average, how many hours do you use a coffee shop as a place to work or study?	Numeric (Continuous)	Average duration spent per working session (in hours).
4.	How much money on average do you spend (in Rupiah) when buying food or beverages while doing tasks at a coffee shop?	Numeric (Continuous)	Average monetary expenditure per visit for food and drinks (in Indonesian Rupiah [IDR]).
5.	How often do you visit the same coffee shop repeatedly (repeat visit) specifically for doing tasks because you feel suitable with its atmosphere?	Categorical	Frequency of returning to the same coffee shop due to a suitable atmosphere.
6.	How many different coffee shops have you visited specifically for doing tasks in the past 1 month?	Numeric (Discrete)	Number of distinct coffee shops visited for work/study in the last month.
7.	Which working area environment do you prefer inside the coffee shop?	Categorical (Nominal)	Preferred seating environment area inside the coffee shop.
8.	When is your favorite time to start working/doing tasks at a coffee shop?	Categorical (Nominal)	Preferred time of day to start working or studying.
9.	How many electronic devices (such as laptops, smartphones, tablets, wireless earphones) do you bring when doing tasks at a coffee shop?	Numeric (Discrete)	Total count of electronic devices brought during a working session.
10.	How often do you feel that your study/work productivity increases when you are at a coffee shop compared to at your room/home?	Categorical	Frequency of perceiving enhanced productivity at coffee shops vs home.
11.	What percentage (%) of your target tasks, on average, do you successfully complete in a single task session at a coffee shop?	Numeric (Continuous)	Average percentage of target tasks successfully completed per session (0–100%).
12.	How often do you feel disturbed by the noise of other visitors who are hanging out while you are focusing on doing tasks?	Categorical	Frequency of experiencing noise disturbance from other patrons.
13.	How often do you experience issues such as slow Wi-Fi connections or full power outlets when intending to do tasks?	Categorical	Frequency of encountering technical facility issues (Wi-Fi/power outlets)
14.	How many colleagues/friends do you usually invite when you want to work or discuss at a coffee shop?	Numeric (Discrete)	Number of friends or colleagues accompanying the respondent.
15.	What is your average number of order items (pcs) during work from a cafe?	Numeric (Discrete)	Average total quantity of items ordered (in pcs) per session.

Based on Table I, this study employs a quantitative experimental framework to investigate and predict workspace comfort levels for patrons working from cafes. The primary dataset comprises 150 valid responses gathered via an online Google Form survey across diverse academic backgrounds, including 30% from the Faculty of Computer Science and 70% from other faculties. The collected variables include numerical attributes like visit duration and weekly frequency, alongside categorical attributes such as seating preferences and facility availability. Finally, the target variable represents the overall perceived workspace comfort level, categorized into discrete classes.

B. Research Ethics

Ethical considerations in this study were addressed by obtaining informed consent from all 150 participating student respondents prior to the survey administration. Participants were fully informed of the research objectives and how their inputs would be utilized. Furthermore, strict measures were maintained to protect the anonymity and confidentiality of all collected data throughout the research process.

C. Data Preprocessing

The data preparation phase ensures optimal dataset quality prior to model training through systematic data cleaning and splitting. This workflow involves filtering out incomplete entries and partitioning the dataset using a stratified 80:20 training-to-testing ratio. Because the dataset contains a hybrid distribution of numerical and categorical attributes, tree-based ensemble frameworks like Decision Trees and Random Forests are leveraged. These algorithms natively process mixed-type variable spaces directly and efficiently, eliminating the need for complex feature scaling, normalization, or encoding transformations.

D. Application of Decision Tree and Random Forest

This research specifically implements and compares two supervised classification algorithms, decision tree and random forest to model and predict coffee shop workspace comfort. Prior to model training, the target variable representing overall customer comfort (originally measured on a Likert scale from 1 to 5) is mapped into three distinct discrete classes to facilitate classification: ratings of 1 to 2 are categorized as *Kurang Nyaman* (Not Comfortable), a rating of 3 as *Nyaman* (Comfortable), and ratings of 4 to 5 as *Sangat Nyaman* (Very Comfortable).

1. This study implements a Decision Tree classification model using Scikit-Learn to establish workspace comfort rules. The preprocessed dataset is partitioned into an 80:20 training-testing split with stratification to ensure reliable evaluation. During training, the algorithm recursively partitions feature spaces such as environmental noise and internet stability to determine optimal split points. To measure node impurity and evaluate the best attributes for splitting, the model utilizes Gini Impurity (1) and Shannon Entropy (2). Furthermore, the selection of optimal splitting features is determined by maximizing the Information Gain (3).

Gini Impurity Formula:

$$Gini(D) = 1 - \sum_{i=1}^c (p_i)^2 \quad (1)$$

Explanation:

D : The current dataset or node being evaluated

c : The total number of classes or target categories

p_i : The proportion or probability of samples belonging to class i within node D

Shannon Entropy Formula:

$$Entropy(D) = - \sum_{i=1}^c p_i \log_2(p_i) \quad (2)$$

Explanation:

D : The current dataset or node being evaluated.

c : The total number of classes or target categories.

p_i : The proportion or probability of samples belonging to class i within node D.

Information Gain Formula:

$$Gain(D, A) = Entropy(D) - \sum_{v \in Values(A)} \frac{|D_v|}{|D|} Entropy(D_v) \quad (3)$$

Explanation:

$Entropy(D)$: The initial entropy of the dataset before splitting based on attribute A.

A : The attribute or feature being tested for splitting.

$Values(A)$: All possible unique values of attribute A.

$|D|$: The total number of sample instances in the parent node.

$|D_v|$: The number of sample instances in the subset branch after splitting by value v of attribute A.

$Entropy(D_v)$: The entropy value of the subset branch D_v

2. This study implements an ensemble Random Forest Classifier using Scikit-Learn to mitigate overfitting and enhance predictive performance. The algorithm applies bootstrap aggregating (bagging) to construct multiple decision trees in parallel on distinct data samples. To ensure tree diversity and reduce correlation, it incorporates randomized feature selection at each node split. Ultimately, predictions are aggregated via majority voting (4), providing superior generalization and robustness for analyzing complex coffee shop workspace attributes.

Majority Voting Formula:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_B(x)\} \quad (4)$$

Explanation:

$\hat{y}$: The final aggregated class prediction produced by the overall Random Forest model.

$\text{mode}\{\dots\}$: The statistical mode function used to identify the most frequently occurring value (majority vote) from a set of predictions.

$h_1(x), h_2(x), \dots, h_B(x)$: The individual class predictions generated by each independent decision tree, ranging from tree 1 to tree.

B : The total number of decision trees (ntree) constructed within the Random Forest ensemble.

x : The input feature vector being evaluated for comfort prediction.

Additionally, the use of the confusion matrix and the visualization of each model's implementation will be presented conditionally based on the resulting accuracy levels to highlight the performance of each method.

E. Model Training

The cleaned dataset undergoes a systematic feature encoding process to transform categorical data into a machine-readable format, expanding the dataset into a 75-column feature space. Specifically, LabelEncoder is applied to the target variable (Comfort Level) and ordinal attributes that possess an intrinsic ranking, such as the frequency of encountering noise disturbances or facility constraints. Conversely, One-Hot Encoding is utilized for nominal categorical variables that lack a natural order, such as preferred working area (indoor and outdoor) and favorite time of day, to prevent the algorithms from misinterpreting arbitrary numbers as hierarchical magnitudes. Following this encoding, the dataset is partitioned using an 80:20 stratified train-test split with a fixed random state. This provides key methodological benefits which are 120 training samples enable pattern learning, while 30 unseen test samples ensure an objective, unbiased evaluation of generalization, and prevent overfitting. Decision Tree and Random Forest models are subsequently trained on this data to capture complex patterns and generate predictions for performance assessment.

F. Model Evaluation

Following the training phase, the performance of both the Decision Tree and Random Forest classifiers is rigorously evaluated using the testing dataset (X_{test} and y_{test}) to measure their generalization capabilities on unseen data. To comprehensively assess how accurately the models categorize customer comfort levels, this study employs four standard classification performance metrics:

1. Accuracy: Measures the overall proportion of correct predictions (both positive and negative) made by the model relative to the total number of test samples (5).

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

Confusion Matrix Notations:

- 1) True Positive (TP): The number of instances belonging to a specific class that are correctly predicted by the model.
- 2) True Negative (TN): The number of instances outside the target class that are correctly identified as belonging to other classes.
- 3) False Positive (FP): The number of instances outside the target class that are incorrectly classified into that class.
- 4) False Negative (FN): The number of instances belonging to a specific class that are incorrectly classified into other classes.
2. Precision: Precision evaluates class-by-class exactness by calculating the ratio of true positive predictions to all predicted positives for each customer comfort category, quantifying how effectively the model minimizes false positives (6).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

3. Recall (Sensitivity): Quantifies the model's completeness by measuring the ratio of true positive predictions to all actual positive instances present in the test set, ensuring minimal false negatives (7).

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

4. F1-Score: Represents the harmonic mean of precision and recall, providing a balanced single-metric evaluation, which is particularly crucial when dealing with subtle variations or skewed distributions across comfort categories (8).

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

Through this evaluation framework, the performance differences between the single-tree architecture and the ensemble bagging approach are quantified, offering a transparent assessment of each model's predictive reliability in analyzing coffee shop workspace dynamics.

G. Analysis

To thoroughly interpret the outcomes of the predictive models, the subsequent analysis phase will focus on examining several core dimensions of the results. First, a comparative evaluation between the single Decision Tree and the ensemble Random Forest model will be conducted to assess how effectively bootstrap aggregating mitigates generalization error and improves overall predictive accuracy. This performance review will specifically examine class-by-class precision across each individual customer comfort category, such as Not Comfortable, Comfortable, and Very Comfortable to quantify the exactness of the models' positive predictions and evaluate how well the algorithms minimize false positives and avoid misclassifying different comfort tiers. Furthermore, an in-depth interpretation of the feature importances extracted from the Random Forest model will evaluate how economic commitments and productivity metrics, including average spending and task duration,

influence workspace comfort compared to environmental attributes. Finally, the white-box hierarchical rules visualized from specific tree estimators will be inspected to identify the exact conditional thresholds separating satisfied remote workers, translating these empirical findings into actionable insights regarding the functional shift of coffee shops into remote workspaces.

H. Flow Chart

Figure 1 illustrates the study's sequential workflow, which includes data collection from 150 respondents, preprocessing, stratified 80:20 dataset partitioning, and training using Decision Tree and Random Forest classifiers. Following model evaluation, the process routes to performance interpretation or hyperparameter optimization prior to final conclusion.

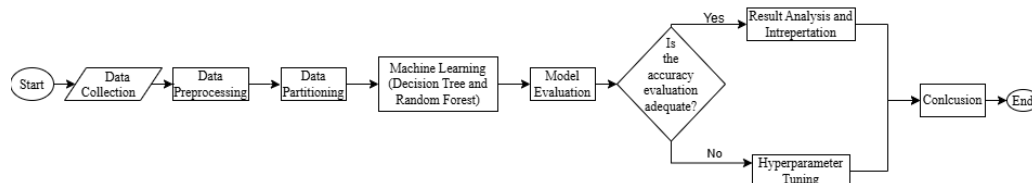


Figure 1. Research methodology workflow

III. RESULTS AND DISCUSSION

A. Description Statistics

To provide an initial overview of the dataset, descriptive statistical analysis was performed on the behavioral traits, environmental preferences, facility constraints, and productivity metrics of the 150 respondents.

TABLE II. Descriptive Statistics Summary of Key Variables

No.	Variable	Dominant Indicator	Value
1.	Target Variable (Comfort Level)	Very Comfortable (Scores 4–5)	65,3%
		Comfortable (Score 3)	24,7%
		Uncomfortable (Scores 1–2)	10%
2.	Preferred Area	Indoor	71,3%
3.	Favorite Time	Night	50%
4.	Facility Constraints	Occasionally	41,3%
		Often	33,3%
5.	Behavioral Metrics	Average Session Duration	4,49 hours
		Average Expenditure (IDR)	Rp28.500
		Average Devices Brought	2,5 unit
6.	Productivity Impact	Task Completion: Very Comfortable Group	74,1%
		Task Completion: Uncomfortable Group	52%

The descriptive results based on Table II, confirm the functional shift of coffee shops into primary workspace alternatives for students, driven by strong indoor (71.3%) and nighttime (50.0%) study preferences. On average, respondents spent 4.49 hours per session with an expenditure of Rp28,500 while carrying 2.5 electronic devices, underscoring their reliance on venue infrastructure. Despite 65.3% of patrons reporting high overall comfort (scores 4-5), technical disruptions remain frequent, with 41.3% encountering slow Wi-Fi or fully occupied power outlets occasionally and 33.3% experiencing them often. Crucially, spatial comfort directly impacts academic output: respondents in the Very Comfortable group successfully completed an average of 74.1% of their targeted tasks, whereas this rate dropped significantly to 52.0% for those in the Uncomfortable category.

B. The Result of Data Preprocessing

The data preprocessing phase ensures optimal dataset quality prior to model training through systematic data cleaning, missing value evaluation, and anomaly detection. The specific procedures, statuses, and descriptions implemented across the dataset are summarized in Table III.

TABLE III. RESUME OF DATA CLEANING, MISSING VALUE, AND OUTLIER

No.	Data Preprocessing Stage	Status	Description
1.	Data Cleaning	Completed	Filtered valid data, removed personal information (names, IDs, contacts), and ensured structural integrity.
2.	Missing Value	Not Found	All 150 respondent records are complete with no missing values (0 empty data).
3.	Outlier Analysis	Found (Retained)	Minor outliers detected (such as duration and cafe variations), but retained as they represent valid behavioral variances among respondents.

The data cleaning phase was conducted to verify the structural consistency of the raw survey dataset. As detailed in Table III, records were filtered to retain valid participants. Personally identifiable information (PII), including full names, student identification numbers, faculty affiliations, and contact details along with auxiliary demographic metadata, were thoroughly removed to protect participant privacy while maintaining strict analytical focus on behavioral workspace variables.

Following the data cleaning process, an exhaustive inspection was executed across all survey attributes to detect any potential missing values or incomplete records. Based on Table III, it confirmed that all 150 respondent records were entirely complete across every feature variable, successfully yielding a missing value count of zero. Consequently, no data imputation techniques, record dropping, or filtering were required at this stage, ensuring that the full sample size remained completely intact to provide a robust, unbiased dataset for subsequent machine learning implementation.

Finally, statistical evaluation for anomalies was conducted on the continuous numerical attributes using the Interquartile Range (IQR) method to assess potential data skewness. Although Table III mentions that minor statistical outliers were identified across specific variables, such as 7 instances in session duration and 11 instances in unique cafe variations. These data points were consciously retained without truncation or removal. This decision was made because these values represented authentic, valid user behaviors reflecting natural variances among respondents rather than data entry errors or structural corruption, thus preserving the complete 150 respondent dataset.

C. The Result of Implementation and Model Evaluation

Following the data preprocessing phase, the curated dataset was partitioned into training and testing subsets using an 80:20 ratio. To predict and classify coffee shop workspace comfort levels across three mapped categories using two supervised machine learning algorithms were implemented which are Decision Tree and Random Forest. The classification performance of each model was evaluated using confusion matrices and standard metrics.

TABLE IV. RESUME OF DATA CLEANING, MISSING VALUE, AND OUTLIER

No.	Evaluation Matrix	Decision Tree			Random Forest		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score
1.	<i>Kurang Nyaman</i>	0.55	0.60	0.57	0.54	0.70	0.61
2.	<i>Nyaman</i>	0.45	0.42	0.43	0.50	0.33	0.40
3.	<i>Sangat Nyaman</i>	0.50	0.50	0.50	0.67	0.75	0.71
4.	Macro Average	0.50	0.51	0.50	0.57	0.59	0.57
5.	Weighted Average	0.50	0.50	0.50	0.56	0.57	0.55
6.	Model Accuracy	0.5000 (50.00%)			0.5667 (56.67%)		

According to Table IV, the evaluation of the Decision Tree model reveals a baseline overall accuracy of 50.0%, indicating that a single hierarchical structure captures fundamental decision rules but struggles with the complex, overlapping boundaries inherent in subjective survey responses. While the model achieves moderate success in identifying lower comfort tiers, it exhibits noticeable confusion when classifying intermediate preferences, reflecting the high-variance nature of single-tree architectures where small shifts in feature thresholds lead to unstable classifications. In contrast, implementing the ensemble Random Forest classifier yields an observable performance enhancement, elevating overall accuracy to 56.67% alongside stronger macro-averaged metrics. By aggregating multiple bootstrap samples and incorporating randomized feature selection, the ensemble approach effectively smooths out individual tree instabilities, demonstrating superior reliability in capturing

extreme sentiment categories. However, both algorithms share a common bottleneck regarding the intermediate comfort class, underscoring that transitional workspace satisfaction levels possess subtle, non-linear feature overlaps that remain challenging to partition cleanly regardless of model complexity.

Reading across Figure 2, the confusion matrix for the Random Forest classifier provides deeper insight into its 56.7% overall accuracy, revealing a strong capacity for identifying extreme comfort tiers while highlighting persistent confusion within intermediate categories. The model correctly identified 7 out of 10 instances of "*Kurang Nyaman*" (with 2 misclassified as "*Nyaman*" and 1 as "*Sangat Nyaman*") and 6 out of 8 instances of "*Sangat Nyaman*" (with 2 misclassified as "*Nyaman*") along its major diagonal. However, the intermediate "*Nyaman*" class served as the primary source of predictive ambiguity, where out of 12 actual instances, only 4 were correctly classified. Notably, 6 actual instances of "*Nyaman*" were misclassified downward into "*Kurang Nyaman*" and 2 instances upward into "*Sangat Nyaman*", indicating that the ensemble model tends to lean toward conservative or lower satisfaction predictions when facing ambiguous, overlapping survey criteria. Meanwhile, misclassifications between the two outer extremes remained minimal (with only 1 instance of "*Kurang Nyaman*" incorrectly predicted as "*Sangat Nyaman*" and 0 vice versa), confirming that while transitional workspace preferences present complex boundary overlaps, the ensemble architecture reliably distinguishes clear-cut sentiment extremes.

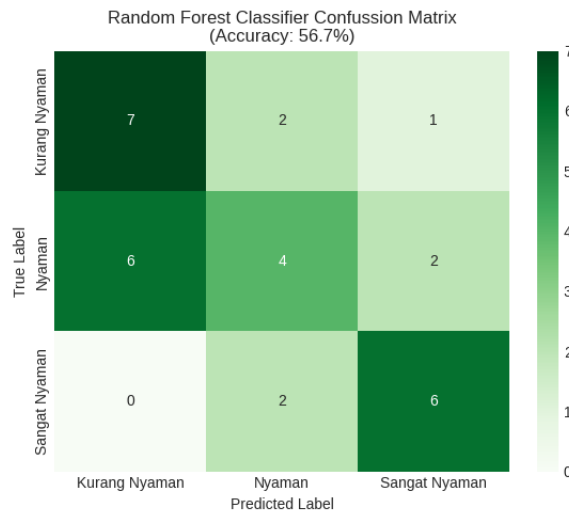


Figure 2. Random Forest Confusion Matrix

Comparative analysis reveals that while the ensemble Random Forest classifier achieved a higher overall accuracy (56.7%) compared to the single Decision Tree (50.0%), both algorithms faced similar limitations in separating the overlapping behavioral traits inherent in the intermediate *Nyaman* category. These findings highlight the complexity of workspace comfort behavior and suggest that future work should incorporate advanced feature engineering or class-balancing techniques to improve inter-class separability. This outcome is primarily influenced by the subjective characteristics of the survey data derived from 150 student respondents, where boundaries between satisfaction levels, especially the transitional "Comfortable" class, are inherently fluid and lack strict mathematical thresholds. Furthermore, as revealed in the confusion matrix, this intermediate class frequently suffers from misclassifications toward the lower tier due to the class overlapping phenomenon, where moderately satisfied respondents exhibit behavioral traits similar to those feeling less comfortable. This condition is compounded by the high dimensionality of the feature space resulting from One-Hot Encoding (expanding to 75 columns), which introduces considerable data sparsity against the moderate sample size.

To understand model dynamics more deeply, evaluation cannot rely solely on accuracy metrics. Comparing Decision Tree and Random Forest uncovers vital trade-offs, where Random Forest demonstrates overall superiority with a Macro Average F1-Score of 0.57 through its bagging mechanism and majority voting, successfully mitigating the high variance inherent to individual trees. Both models achieved peak performance on extreme categories (such as the Very Comfortable class reaching an F1-Score of 0.71), whereas the transitional class received the lowest scores due to behavioral ambiguity. While the single Decision Tree excels in transparent rule extraction, Random Forest trades direct interpretability for greater predictive robustness against unseen data.

Despite these insights, several limitations must be acknowledged. The study relies on a localized

sample of 150 student respondents, predominantly from computer science backgrounds and focuses primarily on visit duration, spending, facilities, and noise disturbances while omitting micro variables like menu variations or interior aesthetics. Consequently, these empirical findings are most accurately generalizable to urban students utilizing coffee shops as communal third spaces. From a practical standpoint, these findings offer actionable strategies for coffee shop management. Management must prioritize core facility reliability by upgrading Wi-Fi bandwidth and expanding electrical power access, implement acoustic zoning to separate social areas from quiet focus study zones, and adjust furniture layouts to support extended study sessions exceeding four hours.

To explain the factors differentiating comfort levels, the decision tree structure on the training data (80% subset, 120 samples) was examined in further detail, as illustrated in Figure 3. At the root node, the Gini impurity value of 0.665 approached the maximum theoretical value for three classes (0.667), indicating that the three comfort categories were mixed almost evenly without a dominant class. The variable representing task completion percentage (*persen_tugas_selesai*) was selected as the primary split with a threshold of 82.5, establishing it as the most influential root predictor. Respondents with a task completion rate of 82.5% or lower comprised the majority branch (69 samples), characterized by high impurity and diverse comfort evaluations. In contrast, the smaller right branch (11 samples) captured respondents with higher task completion rates ($> 82.5\%$), which subsequently branched based on session duration (*durasi_jam*).

Down the primary branch, subsequent splits based on environmental and demographic attributes, task completion percentage (*persen_tugas_selesai*), environmental noise disturbances (*gangguan_kebisingan_Sangat_Sering*), and session duration (*durasi_jam*) consistently emerged as prominent predictors across the constituent trees. By averaging predictions via majority voting, the ensemble effectively reduced the variance and individual overfitting inherent to single decision trees. However, due to the complex and overlapping nature of subjective workspace comfort across the dataset, the ensemble's overall predictive accuracy remained limited on unseen test data. These patterns indicate that while the Random Forest mitigates localized noise better than a standalone tree, future iterations require comprehensive hyperparameter tuning, such as optimizing the number of estimators, maximum tree depth, and minimum samples per split to enhance overall generalizability and robustness.

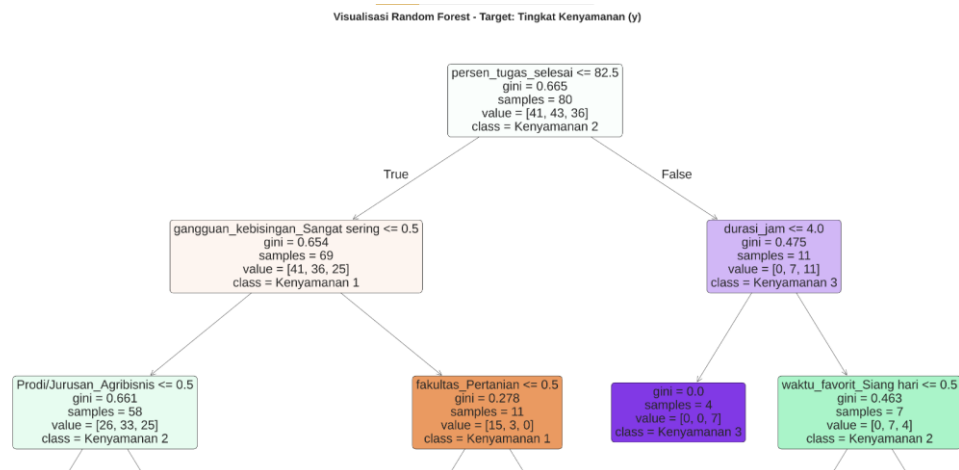


Figure 3. Comfort Level Visualization Using Random Forest

Conclusion

In conclusion, this empirical study demonstrates and validates the modern functional shift of coffee shops, illustrating how traditional social venues have evolved into vital, multi-purpose alternative workspaces and study environments for university students. Through comprehensive exploratory data analysis of 150 valid survey responses, the research reveals that higher perceived workspace comfort levels, predominantly concentrated in the Very Comfortable category (scores 4-5, accounting for 65.3% of respondents) are positively associated with academic and professional productivity, as reflected by average task completion rates rising from 52.0% among Uncomfortable respondents to 64.9% in the Comfortable group and 74.1% in the Very Comfortable group. However, the findings also highlight critical infrastructural bottlenecks, particularly recurring issues with Wi-Fi stability and

power outlet availability, which continue to challenge patrons during extended working sessions.

From a predictive modeling standpoint, the implementation of Decision Tree and Random Forest classifiers yielded modest overall accuracies ranging from 50.0% to 56.7%. This indicates that while tree-based models offer transparent exploratory insights into hierarchical behavioral rules, the current feature set and baseline configurations are not yet fully sufficient for high-precision automated prediction without rigorous hyperparameter tuning, feature refinement, and tree pruning. Ultimately, these insights provide actionable guidance for coffee shop owners and spatial planners, emphasizing that facility optimization, specifically robust internet connectivity and adequate power infrastructure is paramount for meeting the demands of modern remote workers. Furthermore, this study serves as a foundational baseline, opening avenues for future research to incorporate larger datasets, advanced feature engineering, and class-balancing techniques to better capture the complex dynamics of public workspace comfort.

Acknowledgment

The authors gratefully acknowledge the participation and valuable contributions of the 150 respondents who generously completed the survey, making this empirical study on coffee shop workspace comfort levels possible. This research was collaboratively conducted by Fitriany Aisyah Nurvella, Putri Ayu Dita Widya, and Fitri Nur Wahyuni from UPN "Veteran" East Java; Michel Archie Satria Aji Prasetyoo from Istanbul Kent University; Derviola Putri Cahayani from Jember University; and Az Zahra Ailsa Rasyid from Airlangga University. The authors also extend their sincere appreciation to all colleagues and institutions that provided academic support and insights throughout the manuscript preparation process.

REFERENCES

- [1] Liu, M., "In the Post-pandemic and Internet 4.0 era, Cafes' New Operation Mode of Compound Space", *Frontiers in Business, Economics and Management*, 7(1), 24, 2022.
- [2] W. Guo, L. Wang, and L. Caneparo, "Research on the factors that influence and improve the quality of informal learning spaces (ILS) in university campus," *Buildings*, vol. 14, no. 11, Art. no. 3458, Nov. 2024, doi: 10.3390/buildings14113458.
- [3] E. R. Layton and M. Love, "A happy medium: Academic library noise from the perspective of students and librarians," *Codex: J. Louisiana Chapter ACRL*, vol. 6, no. 1, pp. 32–53, 2021.
- [4] D. J. Wibisana, W. Sunarya, and D. K. Santoso, "Servicescape design analysis: The influence of ambient conditions and spatial layout on seating preferences at Bento Café, Surabaya," *Jurnal Arsitektur ZONASI*, vol. 9, no. 3, pp. 549–564, Oct. 2026, doi: 10.17509/jaz.v9i3.97251.
- [5] Y. N. Lukito and A. P. Xenia, "Café as third place and the creation of a unique space of interaction in UI Campus," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 126, no. 1, Art. no. 012208, 2018, doi: 10.1088/1755-1315/126/1/012208.
- [6] Kurnianto, G. W., Aggacitto, & Ardillah, R. T., "Study Time at the Coffee Shop: Exploring Student Motives through a Phenomenological Study", *Indonesian Psychological Research*, 7(1), 43–54, 2025.
- [7] Rompas, A. M., Kusuma, H. E., & Wuisang, C. E. V. (2025). Learning Flow in Café: The Role of Spatial Characteristics. *DIMENSI: Journal of Architecture and Built Environment*, 52(2), 110–120.
- [8] Wicaksono, S. A., Kadek, I., & Nuryana, D., "Perbandingan Klasifikasi Hasil Indeks Kepuasan Masyarakat Menggunakan Algoritma Decision Tree, Random Forest, dan K-Nearest Neighbor", *Emerging Statistics and Data Science Journal*, 3(1), 447, 2024.
- [9] Putra, R. S., Simbolon, H. F. S., Linhar P. A., & Izhari, F., "Perbandingan Algoritma Decision Tree dan Random Forest dalam Klasifikasi Kepuasan Pengguna Sistem Informasi Akademik", *TECHSI - Jurnal Teknik Informatika*, 16(2), 2025.
- [10] Tiftazani, A., Tenriawaru, A., & Rahman, G. A., "Analisis Klasifikasi Kepuasan Mahasiswa terhadap Penyelenggaraan Pelayanan Akademik Menggunakan Algoritma Random Forest", *AnoaTIK: Jurnal Teknologi Informasi dan Komputer*, 2(2), 77–83, 2024.