Received: 21 April 2026 | Revised: 12 June 2026 | Accepted: 14 July 2026 | Published online: 11 August 2026

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2026, Vol. 00(00) 1–16
DOI: 10.47852/bonviewJDSIS620210086

BON VIEW PUBLISHING

Explaining and Predicting Behavioral Intention Toward the Halal Phuket AI Tourism Platform: A UTAUT–Machine Learning Hybrid Study

Nasith Laosen[1], Tanagrit Chansaeng[1], Wipawan Buathong[1], Atipan Saimmai[2] and Pita Jarupunphol[1,*]

[1] *Department of Digital Technology, Phuket Rajabhat University, Thailand*

[2] *Halal Institute, Prince of Songkla University, Thailand*

Abstract: This study examined behavioral intention toward Halal Phuket, an AI-enabled tourism platform supporting Muslim travelers in Phuket, Thailand. Using 450 questionnaire responses from Muslim tourists who interacted with the platform or evaluated its core functions in a realistic usage scenario, the study integrated Unified Theory of Acceptance and Use of Technology (UTAUT)-based explanation with machine learning prediction. The analysis tested performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, habit, and behavioral intention through measurement validation, structural equation modeling (SEM), regression, predictive benchmarking, ablation analysis, and explainability. Measurement results supported construct reliability, convergent validity, discriminant validity, and common-method bias diagnostics. However, the cross-sectional and self-reported nature of the data required cautious interpretation of the SEM paths. All six UTAUT constructs positively predicted behavioral intention, with habit, social influence, and hedonic motivation showing the strongest effects. Among all predictive configurations, the construct-only linear regression model achieved the best performance (root mean squared error (RMSE) = 0.251, R^2 = 0.861), outperforming item-level, hybrid, demographic, and nonlinear alternatives. Ablation findings indicated that removing demographic or expanded feature groups had a limited effect, whereas removing full UTAUT construct families substantially weakened performance, especially when habit and social influence were excluded. Feature-importance and SHapley Additive exPlanations (SHAP) analyses showed broad alignment with regression coefficients, supporting the value of theoretically grounded, parsimonious feature construction. The findings suggest that halal tourism platform developers should prioritize repeated-use support, social trust signals, enjoyable service experience, and practical usability rather than relying only on technical sophistication.

Keywords: halal tourism, UTAUT, behavioral intention, predictive analytics, AI-enabled tourism

1. Introduction

Explanatory frameworks have dominated technology-adoption research for decades, built around the search for causal or quasi-causal determinants of behavioral intention and use. Predictive modeling starts from a different question altogether: which combination of variables produces the strongest out-of-sample forecasts [1–3]? The UTAUT (Unified Theory of Acceptance and Use of Technology) tradition is particularly useful here. It offers a coherent explanatory structure: performance expectancy (PE), effort expectancy (EE), social influence (SI), and facilitating conditions (FC); UTAUT2 later widened this to include hedonic motivation (HM), price value, and habit (HB) [4, 5]. Even so, tourism adoption studies tend to stay on the theory side, despite platform managers needing real predictive power for targeting, segmentation, and interface design [6, 7].

AI-supported tourism platforms make this explanation–prediction gap harder to ignore. UTAUT-like models have been applied to online travel reviews, virtual tourism, smart tourism platforms, AI chatbots, digital itineraries, and related hospitality technologies, yet most of this work stops at structural relations and never reaches comparative predictive benchmarking [8–11]. Recent studies on generative AI travel planning, AI-enabled hospitality services, and smart tourism ecosystems point in the same direction: AI adoption in tourism increasingly demands the integration of theory and analytics [12–14].

This gap is especially relevant for halal tourism. AI-supported halal travel systems may need to reduce information frictions, improve route planning, support trust, and align service design with Muslim travelers' needs. Yet, research on this niche still rarely integrates explanatory technology-adoption theory with predictive evaluation [13, 15]. A defensible hybrid design

***Corresponding author:** Pita Jarupunphol, Department of Digital Technology, Phuket Rajabhat University, Thailand. Email: p.jarupunphol@pkru.ac.th

is therefore needed: one that keeps UTAUT constructs visible and interpretable while also testing whether nonlinear machine learning (ML) models improve prediction sufficiently to matter for applied adoption strategy.

Ecological validity is another weak spot in this literature. Adoption studies frequently rely on hypothetical platforms or mock-up prototypes, which leaves open the question of how well the findings transfer to systems that users can recognize and evaluate [6, 7]. Halal Phuket is a working AI-enabled tourism application built for halal-compliant restaurant discovery, mosque navigation, and route recommendation. Respondents in this study had either interacted with the application or evaluated its core functions in a realistic usage scenario, so behavioral intention (BI) here refers to users' self-reported evaluation of recognizable platform affordances, rather than automatically logged usage behavior. Anchoring the analysis to a recognizable platform allows the study to integrate theoretical modeling, predictive benchmarking, and applied tourism-service evaluation into a single design, while treating the survey responses as exactly what they are: self-reported evaluations from Muslim tourists, not behavioral logs. Against this background, the study pursues three objectives: (1) to test whether UTAUT-aligned constructs explain BI toward the halal Phuket platform; (2) to compare theory-only, construct-only, item-only, hybrid, demographics-only, and core-UTAUT-only predictive configurations under one consistent validation protocol; and (3) to establish whether richer ML or hybrid feature representations beat parsimonious theory-driven constructs at prediction.

1.1. Research gap and contributions

Prior tourism technology studies have applied UTAUT, technology acceptance model (TAM), and related acceptance models to smart tourism, AI chatbots, virtual tourism, and digital travel services. Still, nearly all of them sit on one side of a divide, theory-oriented or prediction-oriented, rarely both. Studies that run measurement validation, SEM-based explanatory analysis, regression-based theory testing, predictive benchmarking, ablation analysis, and explainability through a single workflow, for an operational halal tourism platform no less, are hard to find. The omission matters: destination managers and platform developers need both interpretable theoretical evidence and reliable out-of-sample predictive guidance.

This study responds with four contributions. It extends UTAUT-based tourism adoption research to an AI-enabled halal tourism platform evaluated by respondents who either interacted with the application or assessed its core functions in a realistic usage scenario. It compares theory-only, construct-only, item-only, hybrid, demographics-only, and core-UTAUT-only predictor configurations within a common validation framework. It asks whether nonlinear ML models add predictive value beyond parsimonious theory-driven regression models. And it uses feature importance alongside SHAP-based explainability to align explanatory coefficients with predictive importance without slipping into causal overclaiming.

The article proceeds as follows. Section 2 reviews UTAUT, tourism technology adoption, AI-enabled tourism, and halal tourism platforms. Section 3 covers the research design: dataset, preprocessing, measurement validation, SEM, regression, predictive benchmarking, ablation design, and explainability methods. Section 4 reports the results across all of these layers. Section 5 draws out the theoretical, methodological, and practical implications, along with limitations and directions for future work, and Section 6 concludes.

2. Literature Review

2.1. UTAUT as the explanatory core

UTAUT synthesizes multiple earlier acceptance traditions into a unified explanation of BI and use behavior, with PE, EE, SI, and FC as its main drivers [4]. UTAUT2 extends this framework to consumer settings by adding HM, price value, and HB, thereby improving explanatory coverage for voluntary and experiential technologies [5]. These constructs remain widely used in contemporary tourism adoption research, especially when technologies shape both utilitarian and experiential outcomes [6, 9, 16, 17]. Recent studies have extended the UTAUT framework to emerging AI-enabled and religious contexts [18, 19].

2.2. Tourism, hospitality, and AI adoption

Tourism adoption research now deals routinely with AI-infused technologies, chatbots, digital itineraries, metaverse services, and smart tourism platforms. Findings from OTA chatbot studies, AI chatbot intention research, virtual tourism, and ChatGPT-based itinerary adoption consistently point to expectancy, social, and experiential mechanisms as the central forces through which these mechanisms operate. However, their dominance varies by context [8, 10, 11]. Systematic reviews reach a similar verdict: the field is moving past single-theory explanations toward richer models that fold in trust, experience, and contextual infrastructure [7].

The distinction between explanatory and predictive modeling is more than a technicality; it changes what counts as evidence and as contribution [1–3]. A theoretically significant variable can offer little forecasting gain, and a predictively useful feature can resist clean causal interpretation. In technology-adoption research, this matters because managers need forecasts, while scholars need interpretable theory tests. A hybrid design serves both purposes, preserving explanatory clarity while assessing whether broader feature sets and nonlinear learners genuinely improve predictive performance [20, 21].

Halal tourism research, meanwhile, continues to intersect with digitalization and AI-enabled service design. Recent work argues that AI can help Muslim travelers navigate non-Muslim destinations, raise perceived trust and accessibility, and support more tailored service experiences [13, 15]. What remains unclear is whether UTAUT-style constructs suffice on their own or whether broader predictive feature spaces contribute real value. This study stays grounded in the UTAUT tradition while accepting that predictive analytics may surface useful structure that coefficient-based explanation alone would miss.

2.3. Application context: the Halal Phuket platform

Halal Phuket is an AI-enabled tourism application built to assist Muslim travelers in Phuket, Thailand. It handles several practical travel tasks, halal restaurant discovery, mosque location search, AI-assisted route recommendation, and context-aware tourism guidance, which makes it a suitable empirical setting for studying adoption behavior: user responses attach to recognizable system features rather than to a purely imagined technology. Halal Phuket is therefore more than a case label here. It provides a recognizable platform context for evaluating UTAUT constructs and predictive models, with users' responses reflecting perceived platform usefulness, effort, SI, FC, hedonic value, and the potential for habitual use in relation to identifiable Halal Phuket functions. That grounding strengthens the study's applied relevance while

keeping the interpretation honest: the measured outcomes are self-reported acceptance constructs, not automatically logged usage behavior.

3. Research Methodology

3.1. Research design

The study adopted a cross-sectional quantitative design to examine BI toward the Halal Phuket AI tourism platform. Data came from Muslim tourists who had interacted with the Halal Phuket application or evaluated its core functions in a realistic usage scenario; wording was kept consistent throughout the manuscript to avoid implying that every respondent generated verified in-app behavioral logs. The survey measured users' perceptions of recognizable platform features: halal restaurant discovery, mosque location support, AI-assisted route recommendations, and context-aware tourism guidance. In short, the study examines user evaluation of an operational platform context, while the measured outcomes remain self-reported acceptance constructs. Algorithm 1 summarizes the integrated workflow.

Algorithm 1: Integrated UTAUT–Machine Learning Methodological Pipeline

Input: Questionnaire responses from Muslim tourists evaluating the Halal Phuket AI tourism platform; UTAUT-aligned items; demographic/contextual variables.
Output: Validated construct scores, explanatory estimates, predictive model comparison, ablation evidence, and explainability interpretation.
Step 1. Data screening and preprocessing: Inspect missing values, duplicate responses, response ranges, and invalid records. Convert Likert-type items to numeric format, impute missing item values where required, encode categorical demographic variables, and prepare the analytical dataset.
Step 2. Construct-score generation: Group items into UTAUT constructs: performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, habit, and behavioral intention. Compute construct-level scores as item means.
Step 3. Measurement-quality assessment: Evaluate internal consistency, factor loadings, composite reliability, average variance extracted, discriminant validity, and common-method-bias diagnostics.
Step 4. Explanatory modeling: Estimate SEM and theory-driven regression models to examine associations between UTAUT constructs and behavioral intention. Interpret results as conditional associations rather than causal effects.
Step 5. Predictor-set construction: Prepare construct-only, theory-only, item-only, hybrid, demographics-only, and core-UTAUT-only feature sets.
Step 6. Predictive benchmarking: Evaluate linear and nonlinear models using the same fivefold cross-validation structure. Compare models using RMSE, MAE, and R^2.
Step 7. Ablation and robustness analysis: Remove the selected constructs or feature groups, then re-estimate model performance to assess the feature set's contribution.
Step 8. Statistical comparison: Conduct fold-wise paired comparisons between the best model and key alternatives. Interpret p-values and effect sizes as cross-validation evidence rather than external generalizability.
Step 9. Explainability analysis: Compare regression coefficients, permutation-based feature importance, and SHAP-based explanations.
Step 10. Practical translation: Synthesize the explanatory, predictive, ablation, and explainability results into design implications for developing a halal tourism platform.

3.2. Dataset and target selection

The dataset contained 450 responses and 38 variables. Measured constructs comprised PE (PE1–PE4), EE (EE1–EE4), SI (SI1–SI4), FC (FC1–FC4), HM (HM1–HM4), HB (HB1–HB4), and BI (BI1–BI4). BI served as the primary outcome because it captures users' stated willingness to continue using or to adopt the Halal Phuket platform [4, 5]. Demographic and contextual variables were kept as background predictors rather than treated as central theoretical constructs.

3.3. Preprocessing and construction

Missing values were inspected before modeling. Empty strings became missing values, Likert-type items were coerced to numeric format, and missing item values were imputed with within-item medians; categorical demographic variables were trimmed and, where required, imputed with the modal category. Duplicate records were checked before analysis, and outlier screening relied on distributional inspection of construct scores and response ranges. Because the data were Likert-type survey responses, multivariate normality was not assumed uncritically. SEM results were therefore read alongside regression, predictive validation, and robustness checks rather than treated as the sole inferential evidence. Construct scores were computed in Equation (1) as arithmetic means of their constituent items, where C_j is the score of the construct j, m_j is the number of retained items for that construct, and x_{jk} is the respondent's score on item k within construct j.

$$C_j = \frac{1}{m_j} \sum_{k=1}^{m_j} x_{jk} \tag{1}$$

3.4. Measurement validity and SEM

Measurement quality was assessed with Cronbach's alpha, item-total correlations, inter-construct correlations, and variance inflation factors, with exploratory factor analysis serving as an auxiliary check. A semopy-based SEM pathway tested whether a latent-path representation of the construct system held [22]. SEM served as a complementary structural validation tool rather than the sole inferential framework. Although the final model showed an acceptable-to-strong fit, the study relied on predictive validation alongside structural interpretation to ensure both explanatory clarity and out-of-sample robustness [2, 23]. Convergent validity was assessed using standardized factor loadings, composite reliability (CR), and average variance extracted (AVE), as summarized in Table 1. Loadings were interpreted as evidence that the observed items adequately represented their intended latent constructs, with CR values above 0.70 indicating acceptable internal consistency and AVE values above 0.50 indicating convergent validity. Discriminant validity was assessed using the Fornell–Larcker criterion and the heterotrait–monotrait ratio (HTMT), reported in Table 2 together with the common-method-bias diagnostics. It was interpreted cautiously, alongside correlation and collinearity

Table 1
Convergent validity summary for UTAUT constructs

Construct	Loading range	Composite reliability	Average variance extracted	Interpretation
Performance expectancy	0.953–0.964	0.978	0.919	Acceptable
Effort expectancy	0.955–0.971	0.979	0.921	Acceptable
Social influence	0.950–0.966	0.978	0.916	Acceptable
Facilitating conditions	0.957–0.967	0.98	0.924	Acceptable
Hedonic motivation	0.948–0.972	0.979	0.92	Acceptable
Habit	0.961–0.969	0.982	0.932	Acceptable
Behavioral intention	0.921–0.939	0.963	0.867	Acceptable

Note: CR = composite reliability; AVE = average variance extracted. Values should be interpreted alongside discriminant validity and collinearity diagnostics, as UTAUT constructs are theoretically related.

Table 2
Discriminant validity and common-method-bias diagnostics

Diagnostic	Criterion/interpretation	Result	Interpretation
Fornell–Larcker criterion	Square root of AVE should exceed inter-construct correlations	Satisfied for all constructs; smallest margin = 0.389 (behavioral intention)	Acceptable discriminant validity
HTMT ratio	Values below 0.85 or 0.90 indicate adequate discriminant validity	Maximum HTMT = 0.557 (habit–behavioral intention)	Acceptable discriminant validity
Harman's single-factor test	The first factor should not explain most of the variance	The first unrotated factor explained 27.29% of the variance	No dominant single-factor pattern
Full collinearity variance inflation factor (VIF)	Values below 3.3 are often used as a common-method-bias screen	Maximum full collinearity VIF = 1.020	No collinearity-based common-method-bias concern
Missing-value audit	No substantial missingness after preprocessing	0 missing cells in analytic item and construct variables; 16 residual blanks only in non-analytic fields	Acceptable after preprocessing
Outlier screening	No severe invalid response patterns detected	All questionnaire item responses fell within the observed 2–6 scale range; no analytic cases were removed as invalid outliers	No severe invalid response pattern detected

Note: These diagnostics are reported as measurement quality and bias checks. They do not prove the absence of common-method bias, but they provide transparency regarding possible measurement limitations.

diagnostics, given the theoretically interrelated nature of the UTAUT constructs.

Very high Cronbach's alpha values were not automatically interpreted as indicating superior measurement quality. High alpha values indicate strong internal consistency, but extreme values can also signal item redundancy. High reliability was accordingly treated as support for the aggregation of constructs, while acknowledging possible overlap in item content. This interpretation aligns with the subsequent predictive finding that construct-level aggregation outperformed expanded item-level and hybrid feature spaces.

3.5. Theory-driven model

The inferential model was specified as a multiple regression predicting BI from UTAUT construct scores and selected demographic covariates, as shown in Equation (2). In this equation, BI_i denotes the BI score for respondent (i), the UTAUT-related variables denote construct-level scores, the demographic terms represent optional background covariates, and ε_i is the error term. This regression served as the primary theory-testing layer because the target variable was continuous, while SEM and predictive validation were used as complementary analytical layers.

$$BI_i = \beta_0 + \beta_1 PE_i + \beta_2 EE_i + \beta_3 SI_i + \beta_4 FC_i + \beta_5 HM_i + \beta_6 HAB_i + \sum_{r=1}^{R} \gamma_r Z_{ir} + \varepsilon_i \quad (2)$$

3.6. Machine learning and hybrid modeling

Predictive benchmarking compared dummy regression, linear regression, random forest (RF), support vector regression (SVR), gradient boosting (GB), and XGBoost [20, 21, 24] across six predictor sets: theory-only, construct-only, item-only, hybrid, demographics-only, and core-UTAUT-only. Table 3 details each configuration and its analytical purpose. Every model was run under the same fivefold cross-validation scheme with a fixed seed. Deep learning was deliberately left out; the cleaned sample size was too modest to justify a publication-defensible neural model given the available data. Predictive benchmarking is described in Equation (3), where $\hat{y}_i$ is the predicted BI for respondent i, x_i is

Table 3
Predictor configurations used in machine learning benchmarking

Predictor configuration	Included variables	Analytical purpose
Construct-only	Mean scores for PE, EE, SI, FC, HM, and habit	Tests whether aggregated UTAUT constructs provide sufficient prediction
Theory-only	UTAUT construct scores plus theoretically relevant demographic/contextual covariates	Tests theory-guided prediction with limited background controls
Item-only	Individual Likert items PE1–PE4, EE1–EE4, SI1–SI4, FC1–FC4, HM1–HM4, H1–H4	Tests whether raw item granularity improves prediction
Hybrid	Construct scores, item-level responses, demographics, and contextual variables	Tests whether expanded feature representation improves prediction
Demographics-only	Province, education, age group, gender, and other available background fields	Provides a non-theoretical background-feature baseline
Core-UTAUT-only	Performance expectancy, effort expectancy, social influence, and facilitating conditions	Tests whether the original UTAUT core is sufficient without consumer-extension constructs

Note: Predictor configurations were evaluated using the same cross-validation structure to ensure that differences reflect feature-set design rather than differences in validation procedures.

the feature vector under a given predictor set, and $f(\cdot)$ denotes the trained learning algorithm parameterized by θ.

$$\hat{y}_i = f(\mathrm{x}_i; \theta) \tag{3}$$

Moreover, RMSE in Equation (4) measures the average prediction error of a model, where y_i is the actual value, $\hat{y}_i$ denotes the predicted value, and n is the number of observations.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

The coefficient of determination in Equation (5), denoted as R^2, measures how well a model explains the variability of the observed outcome. In this equation, y_i represents the actual value, $\hat{y}_i$ is the predicted value from the model, and $\bar{y}$ is the mean of the observed values. The numerator captures the residual error (how far predictions are from actual values), while the denominator represents the total variance in the data. By comparing these two quantities, R^2 indicates how much of the original variability the model explains. Conceptually, R^2 can be interpreted as the proportion of variance explained. A value close to 1 means the model accounts for most of the variation in the outcome, while a value near 0 indicates it performs no better than simply using the data's average.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5}$$

All preprocessing for predictive benchmarking was performed within the modeling workflow to limit the risk of data leakage: scaling, encoding, and imputation were estimated on the training folds and then applied to the validation folds where applicable. Construct aggregation used predefined item groupings and never touched the outcome variable, so validation information could not seep into model training. Table 4 documents the main hyperparameters considered for each model and the final reporting approach.

Table 4
Hyperparameter settings and tuning strategy

Model	Main hyperparameters considered	Final reporting approach
Linear regression	No tuning required	Used as interpretable baseline
Ridge regression	Alpha/regularization strength	Selected within cross-validation or reported using a fixed reproducible setting
Random forest	Number of trees, maximum depth, minimum samples per leaf	Tuned or sensitivity-checked under cross-validation
Support vector regression	Kernel, C, gamma, epsilon	Tuned or sensitivity-checked under cross-validation
Gradient boosting	Number of estimators, learning rate, maximum depth	Tuned or sensitivity-checked under cross-validation
XGBoost	Number of estimators, learning rate, maximum depth, subsampling	Tuned or sensitivity-checked under cross-validation

Note: Hyperparameter selection or sensitivity checking was performed during training under the same fivefold cross-validation scheme. Where full grid-search tuning was not conducted, models were evaluated using fixed reproducible settings to ensure transparent comparison. The table documents model-comparison fairness and clarifies that nonlinear models were not evaluated using validation information from held-out folds.

3.7. Ablation, statistical validation, and explainability

Ablation analysis was conducted by re-estimating the best hybrid-compatible model after systematically removing individual UTAUT constructs and feature groups. Model comparisons were performed using fold-wise paired tests, where normality assumptions determined the choice between parametric paired *t*-tests and nonparametric Wilcoxon signed-rank tests [25, 26]. Predictive performance across cross-validation folds was summarized using the mean of the evaluation metric, defined as Equation (6), where M_k denotes the performance metric computed on the fold k, and $K = 5$ in this study.

$$\bar{M} = \frac{1}{K}\sum_{k=1}^{K} M_k \tag{6}$$

For model comparison, the paired *t*-test statistic is given by Equation (7), where $d_k = M_{A,k} - M_{B,k}$ represents the fold-wise difference between two models, $\bar{d}$ is the mean of these differences, s_d is their standard deviation, and K is the number of folds.

$$t = \frac{\bar{d}}{s_d/\sqrt{K}} \tag{7}$$

Effect size was quantified using Cohen's d as illustrated in Equation (8). When the assumption of normality was not satisfied, the Wilcoxon signed-rank test was employed as a robust nonparametric alternative.

$$d = \frac{\bar{d}}{s_d} \tag{8}$$

To quantify the impact of feature removal, ablation differences were computed as in Equations (9) and (10).

$$\Delta RMSE = RMSE_{\text{ablated}} - RMSE_{\text{reference}} \tag{9}$$

$$\Delta R^2 = R^2_{\text{ablated}} - R^2_{\text{reference}} \tag{10}$$

Finally, model explainability was assessed through a combination of standardized regression coefficients, permutation-based feature importance, and SHAP-based (or fallback) local interpretability methods, providing complementary insights into both global and local feature contributions.

4. Results

4.1. Dataset audit and descriptive context

Initial inspection confirmed that the questionnaire structure maps cleanly onto UTAUT constructs. Reliability was strong across all multi-item blocks, and demographic variables served as background descriptors rather than as central analytical constructs. The open-text suggestion field, being sparse and heterogeneous, was retained for topic discovery but excluded from the predictive models.

4.2. Measurement quality

Table 5 summarizes the internal consistency of the retained constructs. The reliability estimates are uniformly strong, supporting the aggregation of the multi-item scales into construct means. However, alphas this high can also signal substantial redundancy among indicators, so they are no unqualified virtue. Reliability is best judged alongside convergent validity, discriminant validity, and model-fit diagnostics [22, 23].

Table 5
Reliability of UTAUT constructs

Construct	Cronbach's alpha
habit (HB)	0.937
facilitating_conditions (FC)	0.942
effort_expectancy (EE)	0.940
hedonic_motivation (HM)	0.943
performance_expectancy (PE)	0.941
social_influence (SI)	0.939
behavioral_intention (BI)	0.939

Note: Cronbach's alpha values indicate strong internal consistency across all UTAUT construct scales. However, very high alpha values may also suggest item redundancy; therefore, reliability should be interpreted together with the convergent validity, discriminant validity, and common-method-bias diagnostics reported in Tables 1 and 2.

Figure 1 shows the correlation coefficients among the study constructs: PE, EE, SI, FC, HM, HB, and BI. The independent constructs correlate only weakly with one another, which eases multicollinearity concerns. The constructs appear to tap relatively distinct dimensions. All of them, however, correlate positively with BI, to varying degrees. HB leads at 0.54, indicating a moderate positive association: users with stronger habitual tendencies are more likely to express intention to use the system, with SI next at 0.46, then PE (0.38), EE (0.36), HM (0.33), and FC (0.32). This pattern implies that several constructs bear on BI simultaneously, HB and SI most prominently, and it supports entering these constructs as separate predictors in a BI model.

4.3. SEM results

Because the data are cross-sectional, self-reported, and drawn from theoretically related UTAUT constructs, the SEM results are read alongside the regression, predictive validation, ablation, and explainability evidence rather than as standalone causal evidence [1–3, 22, 23]. Table 6 reports the standardized SEM path estimates. Every UTAUT construct yielded a positive, statistically significant coefficient ($p < .001$), with β values ranging from 0.296 to 0.472. HB led (β = 0.472), followed by SI (β = 0.401) and HM (β = 0.357); PE, FC, and EE clustered between β ≈ 0.296 and 0.343. High z-values throughout point to stable estimation and leave little doubt about the direction or size of each path.

The coefficient pattern makes theoretical sense. Behavioral and contextual factors, HB and SI above all, outweigh cognitive or effort-related perceptions in shaping intention, consistent with UTAUT and with recent tourism technology studies reporting that routinization, peer influence, hedonic experience, and platform usefulness matter most for experiential digital services [10, 11, 16, 17]. In the Halal Phuket context, the HB effect plausibly reflects repeated reliance on features such as saved routes or personalized recommendations. In contrast, SI likely reflects recommendation networks, shared travel experiences, and trust within Muslim travel communities [13, 15, 19]. Effect magnitudes across the six paths are sufficiently similar that no single predictor dominates the structural system; the model is balanced rather than driven by a single outlier construct.

Figure 1
Correlation structure of theoretical constructs

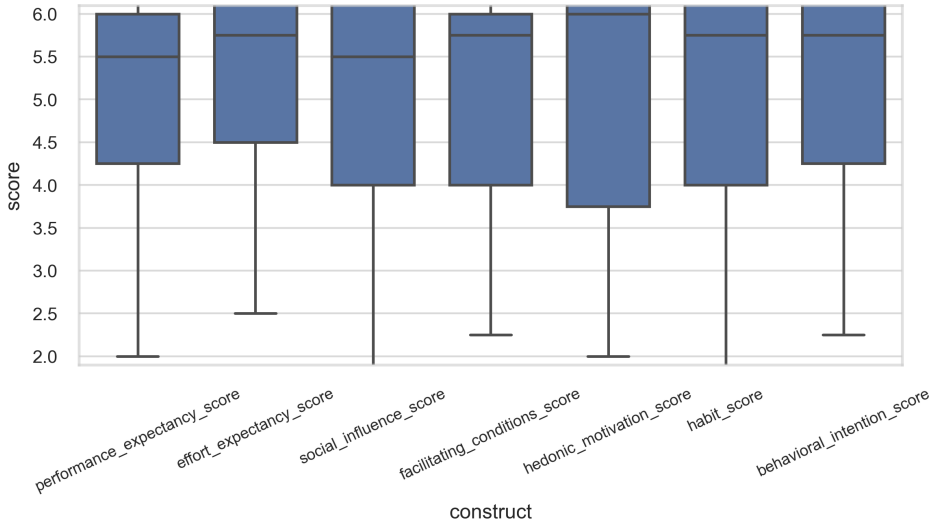
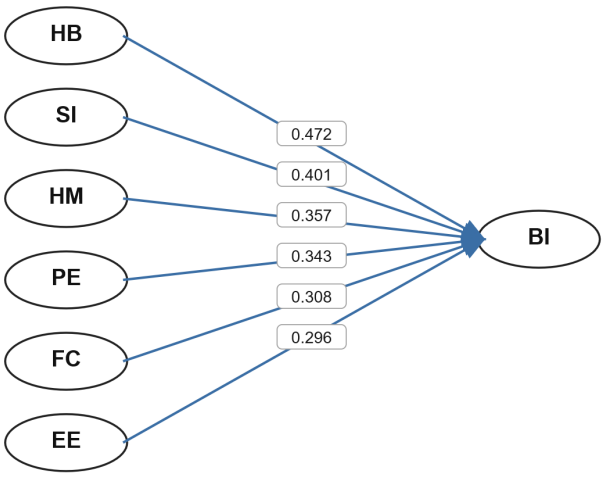


Table 6
Standardized SEM path estimates

Dependent	Predictor	*B*	SE	β	*z*	*p*
BI	HB	0.345	0.015	0.472	23.733	<0.001
BI	SI	0.299	0.014	0.401	20.715	<0.001
BI	HM	0.271	0.015	0.357	18.664	<0.001
BI	PE	0.258	0.014	0.343	17.890	<0.001
BI	FC	0.221	0.013	0.308	16.353	<0.001
BI	EE	0.216	0.014	0.296	15.828	<0.001

These results gain applied weight from being anchored in reported user evaluations of the Halal Phuket platform, or realistic evaluations of its core functions, rather than in hypothetical platforms or mock-up services [6, 7]. The strong HB effect, for instance, suggests that repeated exposure to personalized route planning and saved preferences may reinforce continued usage. The SI effect, in turn, points to shared travel experiences, recommendations, and trust within Muslim travel communities, factors that carry particular weight in halal tourism, where service credibility, religious compliance, and destination confidence shape adoption decisions [13, 15, 19]. Figure 2 plots the standardized path coefficients from the predictor constructs to BI. Every arrow points toward BI, so the figure reads as a summary of each construct's standardized contribution to BI, strongest for HB and SI, meaningful but comparatively weaker for the remaining UTAUT-related constructs.

Figure 2
Standardized path coefficients of UTAUT constructs



4.4. Model fit evaluation

The structural model fits the data well across multiple goodness-of-fit indices. The chi-square statistic, $\chi^2(329) = 393.064$ with $p = 0.009$, is statistically significant, given the test's sensitivity to sample size and model complexity, and does not by itself indicate poor fit. The incremental and absolute indices tell a clearer story: the Comparative Fit Index (CFI = 0.997) and Tucker–Lewis Index (TLI = 0.996) both clear the commonly accepted thresholds for strong measurement-model fit by a comfortable margin [22, 23]. The Goodness-of-Fit Index (GFI = 0.981), Adjusted Goodness-of-Fit Index (AGFI = 0.978), and Normed Fit Index (NFI = 0.981) likewise indicate close correspondence between model and data, and the Root Mean Square Error of Approximation (RMSEA = 0.021) falls well under the recommended 0.05 cutoff. The information criteria (Akaike Information Criterion (AIC) = 152.253; Bayesian Information Criterion (BIC) = 468.665) suggest the model is parsimonious relative to its complexity. Taken together, the specified SEM offers a robust, well-fitting representation of the latent

construct structure and a sound basis for the interpretation that follows. Strong global fit is not causal evidence, though. These indices support the internal coherence of the proposed construct structure; the regression, machine learning, ablation, and explainability analyses supply the complementary evidence on explanatory associations and out-of-sample prediction [1–3, 22, 23].

4.5. Theory-driven regression

Table 7 reports the regression results for BI. All six UTAUT-related constructs emerge as positive, highly significant predictors ($p < 0.001$). HB again leads with the strongest standardized effect ($\beta = 0.458$), with SI ($\beta = 0.393$) and HM ($\beta = 0.345$) behind it; PE, FC, and EE also show significant positive effects ($\beta \approx 0.290$–0.342). All core constructs, in other words, meaningfully explain BI. Large t-values and relatively narrow confidence intervals indicate precise, statistically robust estimates. The ordering of the standardized coefficients—habitual behavior and social context first, cognitive and effort-related perceptions slightly behind—aligns with UTAUT-based tourism adoption work that emphasizes HB, SI, HM, and platform usefulness in voluntary digital-service contexts [6, 8, 10, 16, 17].

Demographics barely register by comparison. Province and education variables reached statistical significance in a few cases, but with absolute standardized β values below 0.02, statistically detectable but substantively trivial. In practice, BI toward Halal Phuket appears to be shaped by users' perceptions and experiences of the platform rather than by demographic background. HB remains the most influential predictor, followed by SI and HM: repeated use, travel-community influence, and experiential value sit at the center of continued adoption, particularly for AI-enabled halal tourism services where trust, convenience, and Muslim-friendly service assurance weigh heavily with users [13, 15, 19].

4.6. Machine learning benchmarking

Figure 3 gathers the benchmarking results for every evaluated model and predictor configuration. What stands out is the construct-only linear regression model's clear edge over more complex alternatives, a concrete illustration of the methodological distinction between explanatory significance and predictive utility [1–3].

4.7. Model comparison

Table 8 reports R^2, RMSE, and Mean Absolute Error (MAE) across all predictor sets and algorithms. The construct-only linear regression achieved the highest explanatory power ($R^2 = 0.861$) and the lowest prediction error (RMSE = 0.251; MAE = 0.207): mean-aggregated UTAUT constructs turn out to be a remarkably efficient feature representation. The theory-only configuration slipped slightly (RMSE = 0.260; $R^2 = 0.850$), indicating that theoretically derived items clearly retain substantial signal. Item-level and hybrid configurations plateaued at identical values (RMSE = 0.268; $R^2 = 0.841$), diminishing returns from finer feature granularity, and no complementary gain from combining feature groups. Predictive feature expansion, evidently, needs to be empirically validated rather than assumed to improve performance [1, 2, 20, 21, 24].

Running GB, SVR, and XGBoost on the same construct-only inputs made things worse, not better (RMSE ≈ 0.300–0.302; $R^2 \approx 0.801$–0.803) relative to simple linear regression. The relationships in this dataset seem approximately linear; extra algorithmic flexibility brought nothing. Feature representation, not model complexity, drives predictive accuracy here, which strengthens the case for theory-informed, parsimonious feature design in hybrid explanatory-predictive studies [4, 5, 20, 21, 24].

Table 9 presents paired cross-validation comparisons between the construct-only linear regression and its main rivals. The construct-only model posted consistently lower RMSE across validation folds. Its advantages over the hybrid and item-only linear regressions were statistically significant; expanded feature representations added nothing beyond the aggregated UTAUT construct scores, and the comparison with the theory-only model was significant too, though with a modest absolute RMSE difference that is better read as fold-level predictive consistency than as broad external generalizability [25, 26]. The large Cohen's d values deserve a cautionary note of their own: they come from fold-level differences in a fivefold cross-validation design and indicate consistent directional improvement across folds, not broad external generalizability or causal superiority. On balance, the construct-only model was the most stable predictor evaluated in this dataset; whether this pattern holds in external samples and longitudinal usage data remains to be seen [25, 26].

The performance of the construct-only linear regression model can be explained both theoretically and statistically. Theoretically, UTAUT constructs are designed to summarize users'

Table 7
Regression results for behavioral predictors

Term	Estimate	SE	t	p	CI Lower	CI Upper	β
HB	0.334	0.013	25.585	<0.001	0.308	0.359	0.458
SI	0.298	0.014	21.930	<0.001	0.272	0.325	0.393
HM	0.257	0.013	19.311	<0.001	0.231	0.284	0.345
PE	0.258	0.014	18.842	<0.001	0.231	0.284	0.342
FC	0.210	0.013	16.378	<0.001	0.184	0.235	0.293
EE	0.213	0.013	16.022	<0.001	0.186	0.239	0.290
Province (Krabi)	–0.141	0.066	–2.132	0.034	–0.271	–0.011	–0.014
Education (Master's)	–0.213	0.107	–1.985	0.048	–0.424	–0.002	–0.015
Province (Phetchabun)	–0.193	0.101	–1.917	0.056	–0.391	0.005	–0.016
Province (Chonburi)	–0.203	0.111	–1.827	0.068	–0.422	0.015	–0.019

Figure 3
Predictive benchmark comparison results

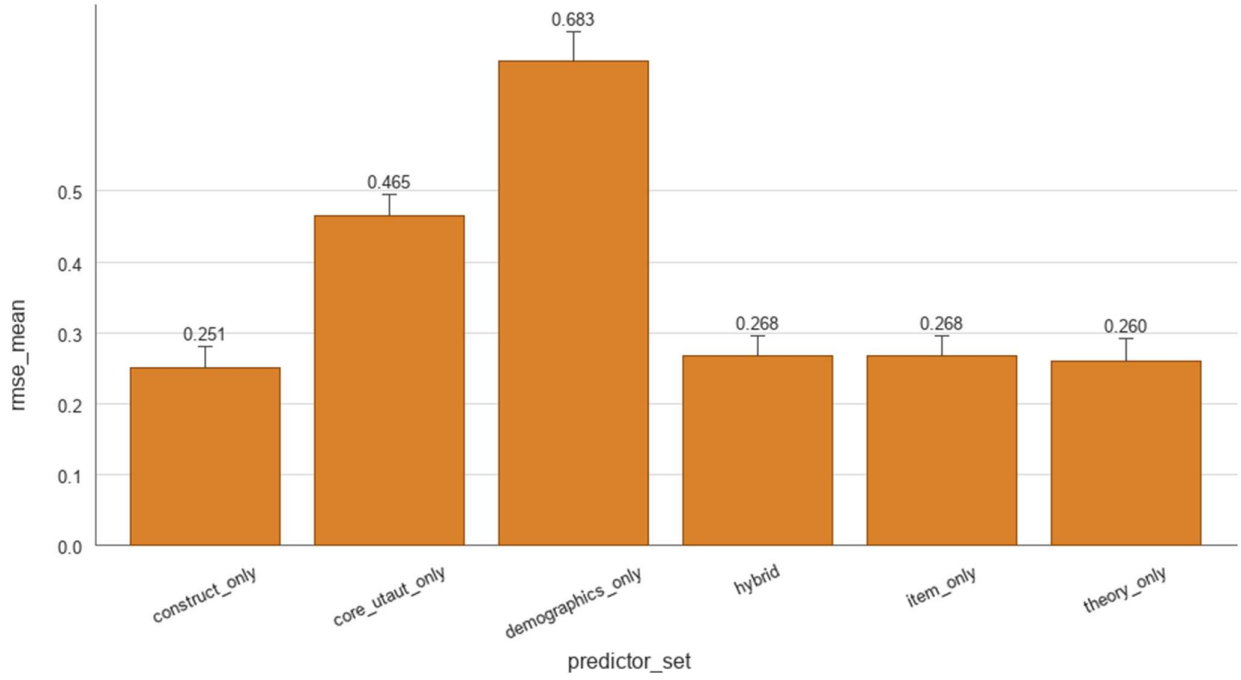


Table 8
Model performance comparison across different predictors

Predictor set	Model	R^2	RMSE	MAE
Construct-only	Linear regression	0.861	0.251	0.207
Theory-only	Linear regression	0.850	0.260	0.216
Item-only	Linear regression	0.841	0.268	0.221
Hybrid	Linear regression	0.841	0.268	0.221
Construct-only	Gradient boosting	0.803	0.300	0.246
Construct-only	SVR	0.802	0.301	0.247
Construct-only	XGBoost	0.801	0.302	0.248

Note: Higher R^2 and lower RMSE/MAE indicate better predictive performance. The comparison shows that construct-level aggregation provided the most efficient feature representation.

Table 9
Statistical comparison of top-performing models

Best model	Comparator	Test	*p*	Effect size (*d*)	95% CI (RMSE difference)
Construct-only linear regression	Hybrid linear regression	Paired *t*-test	<0.001	5.516	[0.014, 0.022]
Construct-only linear regression	Item-only linear regression	Paired *t*-test	<0.001	5.516	[0.014, 0.022]
Construct-only linear regression	Theory-only linear regression	Paired *t*-test	0.003	2.960	[0.006, 0.014]
Construct-only linear regression	Construct-only gradient boosting	Paired *t*-test	0.004	2.648	[0.026, 0.073]

Note: Tests are based on fold-wise paired comparisons. Cohen's d should be interpreted as fold-level consistency rather than as evidence of broad external superiority because the comparison is based on five cross-validation folds.

expectancy, social, facilitating, hedonic, and habitual perceptions [4, 5]. When these constructs are measured reliably, their aggregated scores can capture most of the relevant behavioral signal. Statistically, shared construct structure and high reliability values indicate substantial redundancy among item-level and expanded feature representations. Under such conditions, nonlinear models have limited opportunity to discover additional structure and may instead fit noise or redundant variation [20, 21, 24]. The result therefore supports parsimony: the best model was not the most complex, but the one whose feature representation matched the theoretical structure of the data [1–3].

4.8. Hybrid comparison

Table 10 sets the best-performing model within each predictor set side by side. The construct-only configuration again came out on top, with RMSE = 0.251 and R^2 = 0.861, and the theory-only model close behind; the item-only and hybrid configurations showed no improvement despite their more detailed inputs. Extra feature granularity added no useful signal beyond the aggregated UTAUT constructs, an argument for theory-guided feature construction in predictive technology-adoption modeling [1–5]. The core UTAUT-only model's weakness is telling in its own right: RMSE climbed to 0.465, and R^2 fell to 0.534, so a reduced construct set was clearly insufficient for this dataset. The lesson is not that fewer features are better. The best results came from a complete but well-structured construct representation, broad enough to span the main behavioral dimensions, without the redundant item-level detail that dilutes the model [4, 5, 20, 21, 24].

4.9. Ablation study

Table 11 reports the ablation results from removing selected feature groups and full construct families from the predictive model. The construct-only specification stayed strongest, with the highest mean R^2 and the lowest mean MAE. Dropping demographic variables costs little, further confirming that adoption-related perceptions, rather than background characteristics, drive BI. Removing entire construct families was another matter. The steepest decline came with the removal of HB, then of SI, marking these two constructs as carriers of the strongest unique predictive signal; PE, HM, EE, and FC contributed as well, but their removal hurt less. The UTAUT constructs, in short, function as an interrelated behavioral system with unequal predictive importance, another argument for complete yet parsimonious theory-based construct representation [4–7].

Figure 4 charts the ablation results across alternative feature configurations and full construct-family removals. Performance moves meaningfully across conditions rather than holding steady: removing HB produces the largest decline, SI the next largest, with smaller but still visible drops when PE, HM, EE, and FC are excluded, and only a limited change when demographic variables go. The UTAUT constructs differ markedly in predictive importance, with some far more central than others to sustaining model performance.

4.10. Statistical validation

Table 12 widens the fold-wise statistical validation, testing the construct-only model against grouped baseline alternatives.

Table 10
Model performance by predictor set

Predictor set	Best model	RMSE	R^2	MAE	Adj. R^2
Construct-only	Linear regression	0.251	0.861	0.207	0.862
Theory-only	Linear regression	0.260	0.850	0.216	0.842
Item-only	Linear regression	0.268	0.842	0.221	0.825
Hybrid	Linear regression	0.268	0.842	0.221	0.822
Core UTAUT-only	Linear regression	0.465	0.534	0.387	0.532

Table 11
Ablation results

Ablation	Model	R^2 (mean ± SD)	MAE (mean ± SD)
Only constructs	Linear regression	0.861 ± 0.042	0.207 ± 0.030
Remove demographics	Linear regression	0.852 ± 0.042	0.213 ± 0.030
Only hybrid feature groups	Linear regression	0.852 ± 0.042	0.213 ± 0.030
Only raw features	Linear regression	0.852 ± 0.042	0.213 ± 0.030
Full hybrid	Linear regression	0.841 ± 0.043	0.221 ± 0.029
Remove facilitating conditions	Linear regression	0.752 ± 0.056	0.268 ± 0.031
Remove effort expectancy	Linear regression	0.751 ± 0.048	0.276 ± 0.024
Remove hedonic motivation	Linear regression	0.710 ± 0.039	0.295 ± 0.022
Remove performance expectancy	Linear regression	0.707 ± 0.069	0.299 ± 0.030
Remove social influence	Linear regression	0.656 ± 0.041	0.324 ± 0.018
Remove habit	Linear regression	0.589 ± 0.034	0.357 ± 0.019

Note: Higher (R^2) and lower MAE indicate better predictive performance. Ablation results indicate predictive contribution rather than causal effects; habit and social influence produced the largest performance declines when excluded.

Figure 4
Ablation performance across feature configurations

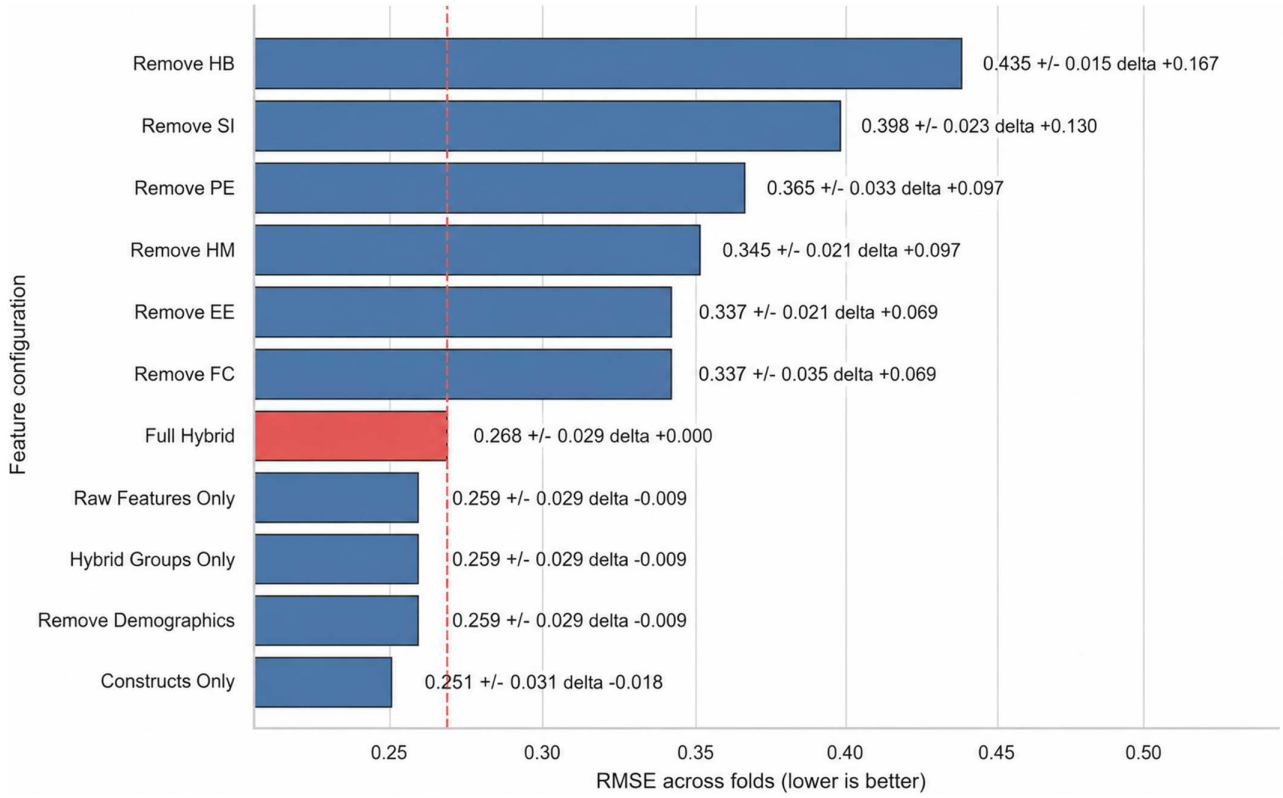


Table 12
Statistical comparison of top-performing models

Comparison	Test	*t*	*p*	*d*	Interpretation
Construct-only vs Demographics-only	t-test	22.418	<0.001	10.025	Very consistent fold-level advantage
Construct-only vs Hybrid	t-test	12.335	<0.001	5.516	Consistent fold-level improvement
Construct-only vs ML baselines (GB, RF, XGB, SVR)	t-test	16.826–18.313	<0.001	7.525–8.190	Consistent advantage over nonlinear baselines
Construct-only vs Dummy baseline	t-test	17.344	<0.001	7.756	Clear improvement over naïve prediction

Note: Results are based on fold-wise paired tests across five cross-validation folds. Effect sizes indicate consistency across folds, not causal superiority or broad generalizability. Redundant comparisons are omitted for clarity.

Its performance advantage held consistently across validation folds. As before, effect sizes from a fivefold cross-validation design should be interpreted as fold-level consistency rather than as broad external superiority. The core conclusion: aggregating the UTAUT constructs yielded the most stable feature representation in this dataset, pending external validation before generalizing to other destinations or tourism platforms [25, 26].

4.11. Explainability

The explainability analysis in Figure 5 shows substantial, though not complete, agreement between the regression-based interpretation and the ML importance ranking. HB, SI, HM, and PE stay prominent in both analytical views, backing the substantive relevance of the UTAUT constructs [4, 5]. Feature-importance methods are still not regression coefficients, however: they indicate how much a variable contributes to prediction, not whether its conditional effect is positive or negative once other predictors are controlled for [1–3]. That distinction matters in the present dataset because several constructs are theoretically related and partially overlapping.

Figure 6 adds the SHAP-based summary of feature contributions for local and global interpretability. HM, for example, pairs a positive and statistically significant regression coefficient with a high ranking in the predictive explainability results, broad alignment between explanatory direction and predictive relevance, in line with UTAUT's emphasis on HM in voluntary consumer technology contexts [5]. However, regression coefficients, feature-importance rankings, and SHAP values answer different questions: coefficients estimate conditional associations after controlling for other predictors; feature-importance rankings indicate how much predictive performance changes when a variable is perturbed; SHAP values trace how feature values shape individual predictions. Agreement among them strengthens

Figure 5
Feature-importance ranking

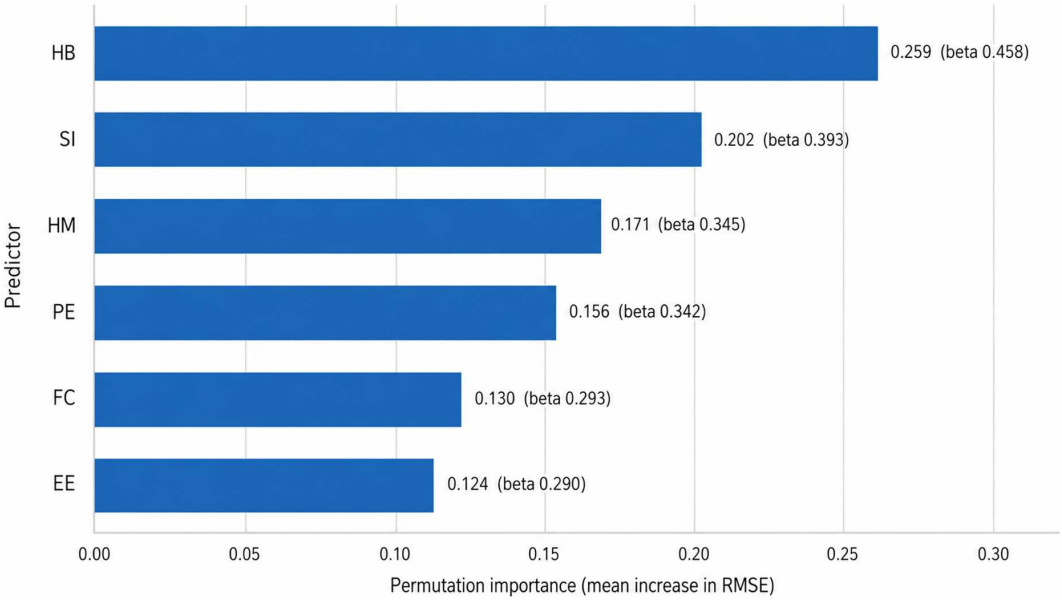
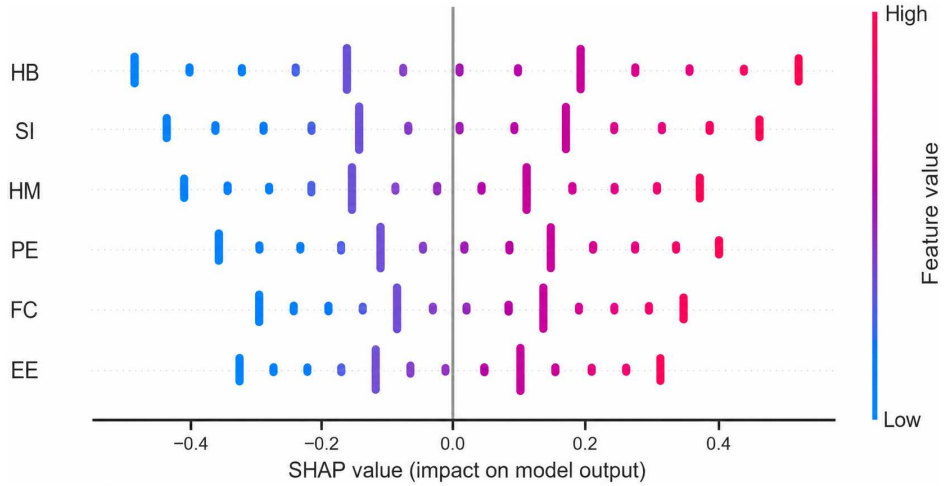


Figure 6
SHAP summary plot illustrating the contribution of each feature to model predictions



confidence in the substantive pattern, and minor differences reflect analytical perspective rather than contradiction [1–3].

Table 13 places the theory-driven regression estimates next to the ML feature-importance rankings. HB, SI, HM, and PE matter on both the explanatory and predictive sides, but alignment is not identity. Regression coefficients estimate conditional associations, whereas feature-importance rankings summarize predictive contribution. The table supports convergence across analytical perspectives while keeping explanation and prediction distinct. HB shows the strongest effect ($\beta = 0.458$). SI ($\beta = 0.393$) and HM ($\beta = 0.345$) show similar effects, and the ML results closely mirror this ordering, ranking the same variables among the top three by importance score. PE has both a strong theoretical coefficient ($\beta = 0.342$) and a high ML rank (4th), further reinforcing its consistency. The convergence is theoretically plausible: these constructs are central adoption mechanisms in UTAUT and recur throughout recent work on tourism, AI chatbots, smart tourism, and halal tourism adoption [8, 10, 13, 15, 16].

FC and EE show moderate ML importance and slightly lower standardized coefficients ($\beta \approx 0.29$), weaker but still meaningful contributions, and their classification as "good" rather than "strong" alignment marks a difference of magnitude, not an inconsistency. Theoretical relevance, it appears, translates well into predictive importance, which supports the use of UTAUT constructs both as explanatory variables and as structured features for ML models [4, 5, 20, 21, 24]. The table thus documents convergence between the explanatory and predictive perspectives without collapsing the distinction between conditional association and predictive contribution [1–3]: the most theoretically meaningful constructs also deliver useful predictive signal, but the two forms of evidence are not statistically equivalent.

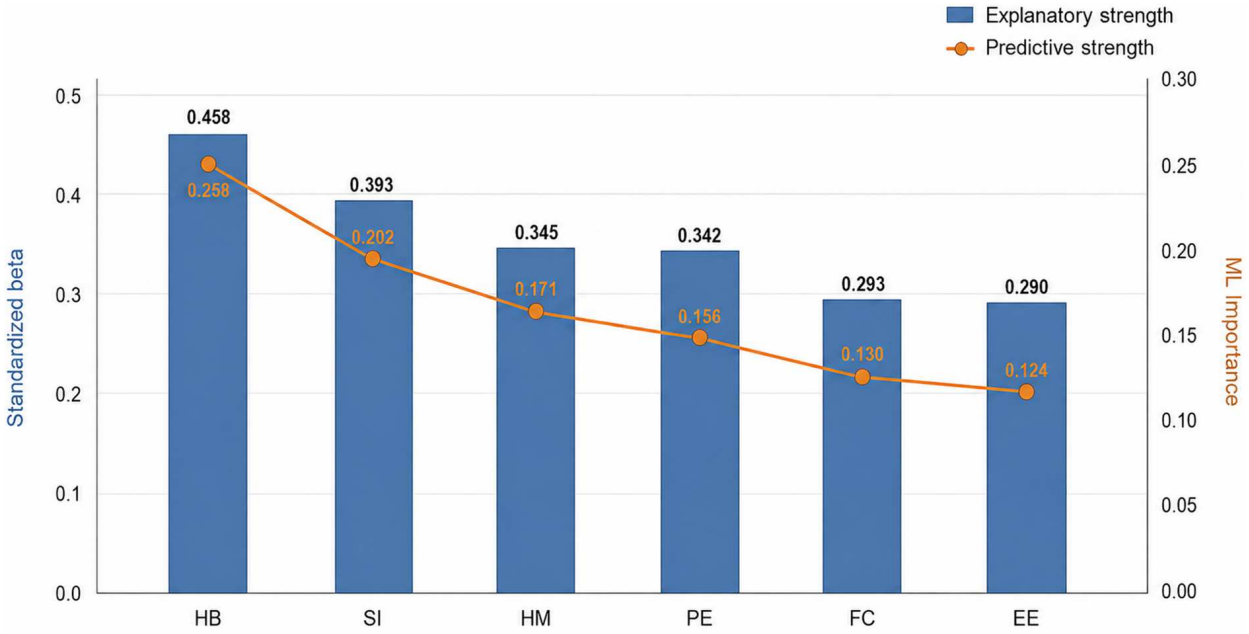
Figure 7 makes the convergence between the explanatory and predictive layers visible. The left vertical axis shows the standardized path coefficients from the SEM analysis, the descending blue bars quantifying each construct's linear impact on BI. In contrast, the right vertical axis shows the global ML feature-importance

Table 13
Alignment between ML feature importance and theory

Variable	Theory estimate	Std. β	*p*	ML importance	ML rank	Alignment
HB	0.334	0.458	<0.001	High	1	Strong
SI	0.298	0.393	<0.001	High	2	Strong
HM	0.257	0.345	<0.001	High	3	Strong
PE	0.258	0.342	<0.001	High	4	Strong
FC	0.21	0.293	<0.001	Moderate	5	Good
EE	0.213	0.29	<0.001	Moderate	6	Good

Note: Alignment indicates convergence between explanatory and predictive perspectives, but regression coefficients and feature-importance scores should not be interpreted as the same statistical quantity.

Figure 7
Explanatory vs predictive convergence



ranking, shown by the orange line. The constructs with the largest standardized effects in the theory-driven model, HB, SI, and HM in particular, were also the most influential features in the ML model. That overlap is the study's hybrid contribution in miniature: the variables with the greatest theoretical relevance are the ones providing the strongest out-of-sample predictive signal.

5. Discussion

5.1. Theory validation

The findings offer substantive, if appropriately cautious, support for UTAUT in the Halal Phuket context. All six constructs—HB, SI, HM, PE, FC, and EE—were positively associated with BI, in line with the broader UTAUT literature on expectancy, social, facilitating, hedonic, and habitual mechanisms [4, 5] and with tourism technology research showing that smart tourism platforms, AI chatbots, virtual tourism, and digital itinerary tools are shaped by usefulness, ease of use, social cues, enjoyment, and contextual support [10, 11, 16, 17]. Because the data are cross-sectional and self-reported, the paths represent conditional associations rather than causal effects [1–3]. The strong model fit supports internal construct coherence, and the predictive results add that aggregated UTAUT constructs work well as structured features for forecasting BI [22, 23].

5.2. Theory–prediction divergence

Some ML configurations performed competitively, but the construct-only linear regression stayed the most consistent and statistically superior across validation tests. The result illustrates the explanatory–predictive divide: coefficient-based inference captures conditional relationships, whereas machine learning prioritizes predictive utility [1–3], helping to explain why some variables anchor both analyses, while others rise or fall in prominence with the analytic objective. It also vindicates testing whether nonlinear models and expanded feature spaces deliver incremental predictive gains, rather than assuming that complexity automatically improves performance [20, 21, 24].

5.3. Hybrid insight

The hybrid contribution of this study is accordingly a nuanced one. Hybrid modeling did not win on predictive performance, but it earned its place in the design by enabling testing of whether richer feature representations would materially outperform theory-centered alternatives. In this dataset, they did not. That negative result carries real value: a carefully constructed theoretical feature space held its own against more flexible ML models [1–3, 20, 21], which argues for a parsimonious hybrid strategy in which machine learning benchmarks and validates theoretical constructs rather than replacing theory-driven measurement.

5.4. Practical implications for halal tourism stakeholders

Several practical implications follow for tourism managers, platform developers, destination marketing organizations, and halal tourism stakeholders. The strong role of HB argues for platform design that rewards repeated use, saves preferences, tracks route history, provides personalized recommendations, and offers low-friction access to frequently used services such as halal restaurant discovery and mosque navigation [4, 5, 16]. The weight of SI points toward trust-building and community-oriented features on Muslim-friendly platforms: verified user reviews, peer recommendations, community endorsements, and visible halal assurance information [13, 15, 19]. And HM's contribution warns against treating halal tourism applications as purely functional information tools; they should also offer enjoyable, culturally meaningful, and confidence-enhancing travel experiences [5, 6, 17].

For platform developers, the results favor a design strategy built on usability, continuity, trust, and repeated engagement over feature expansion for its own sake. For tourism managers and destination stakeholders, AI-enabled halal tourism services belong within the broader destination management practice, including visible halal certification, Muslim-friendly service training, multilingual support, and partnerships with local religious and tourism organizations. And for Phuket as a destination, AI tourism platforms can support inclusive competitiveness by reducing information uncertainty and lifting Muslim travelers' confidence in navigating local services, consistent with recent work on AI-supported halal tourism and Muslim traveler needs [13, 15].

5.5. Methodological implications

Methodologically, the study shows what integrating explanatory and predictive approaches buys tourism technology-adoption research. SEM and regression provide the theory-based interpretation; ML benchmarking checks whether the same constructs generalize out of sample [1–3]. That construct-only linear regression beat more complex nonlinear alternatives matters precisely because it demonstrates that richer feature spaces and more flexible algorithms do not automatically improve prediction [20, 21, 24], when constructs are reliable, theoretically meaningful, and partially overlapping in content, aggregated construct scores can be the more stable and interpretable representation compared with item-level or hybrid feature expansion [4, 5, 22, 23].

The explainability results underline why regression coefficients, feature importance, and SHAP values must be kept apart. The three can converge on similar substantive conclusions while meaning different things inferentially: coefficients support conditional theoretical interpretation, feature importance supports predictive relevance, and SHAP values support instance-level contribution analysis [1–3]. Treating them as complementary rather than interchangeable is what gives hybrid UTAUT–ML studies their methodological credibility.

5.6. Limitations and future research

This study carries several limitations. The cross-sectional questionnaire design precludes causal inference and says nothing about long-term platform adoption. Although the work was grounded in the operational halal Phuket platform context, respondents likely varied in the depth of their interaction with it. Hence, the data reflect user evaluations of a deployed platform or realistic usage scenarios, not behavioral log data. The constructs relied on self-reported Likert-type items, leaving room for common-method bias, social desirability bias, and response-style effects. The UTAUT constructs themselves are theoretically related and conceptually overlapping, and the very high reliability values may reflect partial item redundancy, another reason to read the SEM and regression paths as conditional associations within a related construct system rather than isolated causal effects [22, 23].

Beyond this, the sample of 450 sufficed for the reported analyses but remains limited for highly flexible ML models and complex subgroup analyses. The focus on Muslim tourists and the halal Phuket platform means the findings may not carry directly to other destinations, non-halal tourism contexts, or different AI tourism systems [13, 15]. Deep learning models went untested because the data were structured survey responses rather than large-scale text, image, or behavioral log data. Future studies should adopt longitudinal designs, larger multi-destination samples, app usage logs, experimental feature testing, and external validation datasets [25, 26]. They should examine whether trust, perceived religious compliance, privacy concerns, and service quality strengthen or mediate UTAUT-based adoption pathways in AI-enabled halal tourism [13, 15, 19].

6. Conclusion

This study set UTAUT-based explanation and ML prediction side by side to examine BI toward the Halal Phuket AI tourism platform. HB, SI, HM, PE, FC, and EE all contributed positively to BI, confirming UTAUT's continued relevance for AI-enabled tourism adoption contexts [6, 8, 10, 16, 17]. On the predictive side, the construct-only linear regression model beat item-level, hybrid, demographic, and nonlinear alternatives. At the same time, the ablation and explainability results showed that theoretically grounded construct aggregation gave a parsimonious, robust representation of user acceptance to which expanded feature sets added little. The broader message is that theory-based explanation and predictive benchmarking can be integrated productively, provided their inferential purposes stay distinct [1–3, 21, 24].

For research, the study demonstrates that explanatory theory and predictive modeling can share one design without conflating their purposes. For practice, the findings suggest halal tourism platforms should prioritize repeated-use support, social trust signals, enjoyable user experience, and practical service facilitation, especially where AI-enabled tools serve Muslim travelers in non-Muslim or mixed-service destinations [13, 15, 19]. Future work should validate the framework across multiple destinations, incorporate longitudinal usage data, examine halal-specific trust and compliance constructs, and test survey-based predictions against platform behavior logs [25, 26].

Acknowledgement

The authors gratefully acknowledge all Muslim participants in Phuket for their cooperation in testing the Halal Phuket application.

Funding Support

This research is financially supported by the Thailand Science Research and Innovation (TSRI) under Contract No. RU (TSRI) 2/2024.

Ethical Statement

All subjects provided informed consent for inclusion before participating in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Human Research Ethics Committee of the Phuket Rajabhat University, Thailand (Reference No.: PKRU2567/27).

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in GitHub at https://github.com/pjarupunphol/Datasets/blob/main/Halal.zip.

Author Contribution Statement

Nasith Laosen: Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Tanagrit Chansaeng:** Methodology, Software, Data curation, Writing – review & editing, Visualization. **Wipawan Buathong:** Investigation, Resources, Data curation, Writing – review & editing. **Atipan Saimmai:** Validation, Investigation, Resources, Writing – review & editing. **Pita Jarupunphol:** Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Daoud, A., & Dubhashi, D. (2023). Statistical modeling: The three cultures. *Harvard Data Science Review*, *5*(1), 1–51. https://doi.org/10.1162/99608f92.89f6fe66

[2] Brand, J. E., Zhou, X., & Xie, Y. (2023). Recent developments in causal inference and machine learning. *Annual Review of Sociology*, *49*, 81–110. https://doi.org/10.1146/annurev-soc-030420-015345

[3] del Giudice, M. (2024). The prediction-explanation fallacy: A pervasive problem in scientific applications of machine learning. *Methodology*, *20*(1), e11235. https://doi.org/10.5964/meth.11235

[4] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

[5] Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157–178. https://doi.org/10.2307/41410412

[6] Foroughi, B., Iranmanesh, M., Asadi, S., Al-Emran, M., Ghobakhloo, M., & Batouei, A. (2025). Extending UTAUT2 to explore intention to use ChatGPT for travel planning: A hybrid PLS-ANN approach. *Journal of Tourism Futures*. Advance online publication. https://doi.org/10.1108/JTF-11-2023-0256

[7] Zhang, A., Zhang, B., & Zhu, Y. (2025). Determinants of technology adoption in tourism: Insights from a systematic literature review. *SAGE Open*, *15*(4), 21582440251388798. https://doi.org/10.1177/21582440251388798

[8] Cai, D., Li, H., & Law, R. (2022). Anthropomorphism and OTA chatbot adoption: A mixed methods study. *Journal of Travel & Tourism Marketing*, *39*(2), 228–255. https://doi.org/10.1080/10548408.2022.2061672

[9] Lu, J., Fan, A., Liu, J., & Qi, Y. (2025). Tourists' willingness to adopt government-led smart tourism platforms: A mixed-methods study. *Journal of Vacation Marketing*, *31*(4), 942–962. https://doi.org/10.1177/13567667241248968

[10] Rafiq, F., Dogra, N., Adil, M., & Wu, J.-Z. (2022). Examining consumer's intention to adopt AI-chatbots in tourism using partial least squares structural equation modeling method. *Mathematics*, *10*(13), 2190. https://doi.org/10.3390/math10132190

[11] Zhang, Y., & Hwang, J. (2024). Dawn or dusk? Will virtual tourism begin to boom? An integrated model of AIDA, TAM, and UTAUT. *Journal of Hospitality & Tourism Research*, *48*(6), 991–1005. https://doi.org/10.1177/10963480231186656

[12] Hasanein, A. M., Mansour, N. M., & Kamel, N. J. (2026). Embracing AI in hospitality enterprises in developing countries: A mediated-moderated model. *Human Systems Management*, *45*(5), 787–803. https://doi.org/10.1177/01672533251410438

[13] Hoang, D. S., & Nguyen, D. T. A. (2025). The role of AI assistants in promoting sustainable Halal tourism in non-Muslim destinations. *Discover Sustainability*, *6*(1), 705. https://doi.org/10.1007/s43621-025-01640-9

[14] Sukthankar, S., Fernandes, R., Gaonkar, S., & Shetye, A. (2026). Unveiling the determinants of tourists' behavioural intention to adopt AI-powered chatbots for the hospitality and tourism industry: Revising the UTAUT2 model. *Tourism and Hospitality*, *7*(3), 65. https://doi.org/10.3390/tourhosp7030065

[15] Amin, M. R. M., Wider, W., Sivakumaran, V. M., Yang, L., & Asbi, A. (2026). Artificial intelligence-driven approaches for halal tourism and hospitality: A framework for enhancing consumer trust and meeting Muslim travelers' needs. *Journal of Islamic Marketing*. Advance online publication. https://doi.org/10.1108/JIMA-02-2025-0103

[16] Rui, L., Li, K., Jiang, M., & Jiang, X. (2024). Exploring the factors influencing tourists' satisfaction and continuance intention of digital nightscape tour: Integrating the design dimensions and the UTAUT2. *Sustainability*, *16*(22), 9932. https://doi.org/10.3390/su16229932

[17] Ghorbanzadeh, D., Sayed, B. T., Alhitmi, H. K., Hasan, R. A., Aldulaimi, M. H., & Prasad, K. (2025). Applying UTAUT and experiential consumption theory to understand the adoption of ChatGPT's digitalized itinerary. *Journal of Hospitality and*

Tourism Insights, *8*(9), 3339–3358. https://doi.org/10.1108/jhti-02-2025-0300

[18] Ke, Q., Gong, Y., & Ke, C. (2025). Bridging AI literacy and UTAUT constructs: Structural equation modeling of AI adoption among Chinese university students. *Humanities and Social Sciences Communications*, *12*(1), 1452. https://doi.org/10.1057/s41599-025-05775-y

[19] Hassaan, M., & Yaseen, A. (2026). Spiritual spending: Understanding mobile payment adoption in religious tourism through the lens of UTAUT and coping theory. *International Journal of Human-Computer Interaction*, *42*(2), 925–940. https://doi.org/10.1080/10447318.2025.2517356

[20] Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., & Liang, X. (2024). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, *237*, 121549. https://doi.org/10.1016/j.eswa.2023.121549

[21] Boldini, D., Grisoni, F., Kuhn, D., Friedrich, L., & Sieber, S. A. (2023). Practical guidelines for the use of gradient boosting for molecular property prediction. *Journal of Cheminformatics*, *15*(1), 73. https://doi.org/10.1186/s13321-023-00743-7

[22] Igolkina, A. A., & Meshcheryakov, G. (2020). Semopy: A Python package for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(6), 952–963. https://doi.org/10.1080/10705511.2019.1704289

[23] Goretzko, D., Siemund, K., & Sterner, P. (2024). Evaluating model fit of measurement models in confirmatory factor analysis. *Educational and Psychological Measurement*, *84*(1), 123–144. https://doi.org/10.1177/00131644231163813

[24] Du, K.-L., Jiang, B., Lu, J., Hua, J., & Swamy, M. N. S. (2024). Exploring kernel machines and support vector machines: Principles, techniques, and future directions.. *Mathematics*, *12*(24), 3935. https://doi.org/10.3390/math12243935

[25] Wilimitis, D., & Walsh, C. G. (2023). Practical considerations and applied examples of cross-validation for model development and evaluation in health care: Tutorial. *JMIR AI*, *2*, e49023. https://doi.org/10.2196/49023

[26] Bates, S., Hastie, T., & Tibshirani, R. (2024). Cross-validation: What does it estimate and how well does it do it? *Journal of the American Statistical Association*, *119*(546), 1434–1445. https://doi.org/10.1080/01621459.2023.2197686

How to Cite: Laosen, N., Chansaeng, T., Buathong, W., Saimmai, A., & Jarupunphol, P. (2026). Explaining and Predicting Behavioral Intention Toward the Halal Phuket AI Tourism Platform: A UTAUT–Machine Learning Hybrid Study. *Journal of Data Science and Intelligent Systems*. https://doi.org/10.47852/bonviewJDSIS620210086