Volume XI Issue 2 Year 2026 | Page 573-582 | ISSN: 2527-9866
Received: 29-04-2026 | *Revised: 16-05-2026* | *Accepted: 27-05-2026*

Bridging Theory and Prediction: A Hybrid Explainable SEM–Machine Learning Approach to Consumer Purchase Intention

Sussy Susanti[1], Patah Herwanto[2], Henny Utarsih[3]
[1,2,3]Ekuitas University, Bandung, Jawa Barat, Indonesia, 40124
e-mail: sussy.susanti@ekuitas.ac.id[1], pherwanto@ekuitas.ac.id[2], Henny.utarsih@gmail.com[3]
*Correspondence: *sussy.susanti@ekuitas.ac.id*

Abstract: *The growing use of Instagram as a visual and interactive marketing platform has intensified scholarly interest in how social media content shapes consumer purchase intention. However, most prior studies have relied either on theory-driven Structural Equation Modeling (SEM) or data-driven machine learning, with limited integration between causal explanation, predictive evaluation, and model interpretability. This study addresses this methodological gap by proposing a hybrid explainable SEM–machine learning framework that combines PLS-SEM, XGBoost, and SHAP to examine the relationship between social media content, brand image, and purchase intention. Data were collected from 500 Indonesian Instagram users exposed to fashion and lifestyle brand-related content. The PLS-SEM results show that social media content significantly affects brand image ($\beta = 0.581$, $p < 0.001$), while brand image significantly influences purchase intention ($\beta = 0.511$, $p < 0.001$). Brand image also significantly mediates the relationship between social media content and purchase intention, with a significant indirect effect ($\beta = 0.297$; 95% BC-CI: 0.241–0.356). In the predictive stage, Linear Regression and tuned XGBoost demonstrated stable generalization, with test R^2 values of 0.288 and 0.277, respectively, while Random Forest showed overfitting with a negative test R^2. SHAP analysis revealed that brand image was the strongest predictive feature (mean |SHAP| = 0.302), followed by social media content (0.268), indicating that brand image plays a more prominent role in forecasting purchase intention. The findings contribute theoretically by reinforcing brand image as a key mediating mechanism, methodologically by integrating validated latent constructs into explainable machine learning, and practically by offering digital marketers a dual-lens approach that combines structural explanation with predictive importance.*

Keywords: *social media content, brand image, purchase intention, PLS-SEM, XGBoost, explainable AI*

1. Introduction

Instagram, with over two billion monthly active users as of 2023, has evolved beyond a social platform into a primary channel for brand communication built around visual storytelling. Brands invest heavily in content spanning informational, entertainment, visual, interactive, and creative formats on the assumption that well-crafted content builds brand associations that ultimately drive purchase behavior [1], [2]. Despite the practical significance of this assumption, its underlying mechanisms particularly the pathway through brand image have not been examined with sufficient methodological depth in the literature. Different content types operate through distinct psychological mechanisms. Informational posts reduce decision uncertainty [3], entertainment content strengthens affective attachment through repeated engagement [3], and interactive formats deepen consumer–brand relationships beyond passive exposure [4]. Collectively, these dimensions shape how consumers form and hold brand image defined as the set of associations they link to a brand in memory [5]. A positive, coherent brand image lowers perceived risk and translates directly into purchase intention [6], yet most studies treat this mediation pathway in simplified terms, often testing direct content–intention effects without examining whether the full indirect chain holds across analytical paradigms.

Structural Equation Modeling (SEM) dominates this literature and confirms that social media engagement shapes brand attitude and behavioral intention [7]. Its limitations, however, are well known: linearity assumptions, sample-specific explanatory power, and an inability to evaluate predictive generalizability to new data [7], [8]. Machine learning addresses the predictive gap XGBoost captures non-linear relationships and generalizes through cross-validation [9] but sacrifices theoretical structure, making findings difficult to interpret within behavioral frameworks [10].

Recent hybrid attempts have not fully resolved this tension. Ayad [11] found that most SEM–ML studies treat the two methods as sequential pipelines rather than conceptually integrated frameworks. Ferdinand and Wijaya [12] demonstrated PLS-SEM latent scores as ML inputs but omitted systematic comparison of SEM coefficients with feature importance values. Li et al. [13] used SEM for dimensionality reduction without applying explainability methods to interpret the resulting model. None of these studies employs a framework that simultaneously validates theoretical relationships, evaluates out-of-sample predictive performance, and uses Explainable AI (XAI) to align outputs from both paradigms. Empirically, this study focuses on Indonesian Instagram users who are exposed to brand-related content in fashion and lifestyle product categories, where visual storytelling, entertainment value, interactivity, and brand image play central roles in shaping consumer purchase intention. This context is particularly relevant because fashion and lifestyle brands rely heavily on visually appealing and emotionally engaging content to build brand associations and influence consumer decision-making.

PLS-SEM is employed to test the theoretical model, particularly the mediating role of brand image between social media content and purchase intention. XGBoost is then used to assess the predictive structure of the same theory-derived constructs, while SHAP quantifies the marginal contribution of each feature, allowing a structured comparison between SEM path coefficients and predictive importance values. Three hypotheses are examined: (H1) social media content positively affects brand image; (H2) brand image positively affects purchase intention; and (H3) brand image mediates the relationship between social media content and purchase intention. The contributions of this study are threefold: theoretically, it reinforces brand image as a traceable explanatory mediator; methodologically, it introduces PLS-SEM as a construct-validation step for machine learning; and practically, it provides a dual-lens framework combining structural explanation and predictive importance for digital marketing decision-making.

2. Methods

2.1 Research Design

A quantitative cross-sectional design combined PLS-SEM with supervised machine learning. PLS-SEM was selected for its robustness to non-normality and its suitability for composite-based attitudinal constructs [8]. Rather than treating the two methods as parallel analyses, this study uses PLS-SEM both as a theory-testing instrument and as a feature-engineering step: the latent variable scores generated by PLS-SEM provide psychometrically validated construct scores; SMC and BI were used as input features, while PI served as the prediction target in the machine learning stage. This design preserves theoretical grounding while avoiding the ambiguity between predictor constructs and outcome constructs. SHAP then provides interpretability by decomposing XGBoost predictions into construct-level contributions, enabling a structured cross-method comparison between SEM path coefficients and machine learning feature importance [10], [12].

2.2 Sampling and Data Collection

Data were collected from Indonesian Instagram users (October–November 2024) via purposive sampling. Eligibility required: (1) active Instagram use at least three times weekly; (2) following at least one brand account; (3) encountering brand-related content within the preceding 30 days; and (4) having made at least one online purchase in the preceding three months. Of 543 questionnaires distributed, 500 were retained after removing responses with missing data or straight-line patterns meeting the minimum threshold for complex PLS-SEM models [8]. Respondent characteristics are summarized in Table 1.

Table 1. Respondent Profile (n = 500)

Category	Sub-category	n (%)
Gender	Female	287 (57.4%)
	Male	213 (42.6%)
Age	18–24 years	198 (39.6%)
	25–34 years	224 (44.8%)
	35–44 years	78 (15.6%)
Education	Senior high school	142 (28.4%)
	Undergraduate (S1)	278 (55.6%)
	Postgraduate	80 (16.0%)
Instagram use/day	< 1 hour	63 (12.6%)
	1–3 hours	219 (43.8%)
	> 3 hours	218 (43.6%)
Online purchase (3 mo)	Never	47 (9.4%)
	1–3 times	241 (48.2%)
	≥ 4 times	212 (42.4%)

Note. Percentages may not sum to 100% due to rounding.

2.3 Measurement Instruments and Common Method Bias

Three constructs were measured: Social Media Content (SMC, 5 items), Brand Image (BI, 8 items), and Purchase Intention (PI, 4 items). SMC items informational (IG_INF), visual (IG_VIS), entertainment (IG_ENT), interactivity (IG_INT), and creativity (IG_CRE) were adapted from Tafesse and Wien [3] and Yadav and Rahman [2] and treated as reflective indicators of an overarching latent content quality construct. BI items capturing brand associations, perceived quality, and emotional appeal were adapted from Keller and Swaminathan [5] and Hsu et al. [6]; PI items reflecting behavioral readiness to purchase from [6]. All items used a five-point Likert scale; Indonesian translations were back-translated by a bilingual researcher for semantic equivalence. Representative items are in Table 2.

Table 2. Measurement Instruments Representative Items

Construct	Representative Item	Ref.	Scale
SMC (5 items)	The content is creative and distinctive from other brands (IG_CRE) The content entertains me and keeps me engaged (IG_ENT) Provides useful product information (IG_INF)	[2],[3]	5-pt
Brand Image (8 items)	This brand has a strong and positive identity in my mind (BI_1) I associate this brand with quality and reliability (BI_4)	[5],[6]	5-pt
Purchase Intention (4)	I intend to purchase from this brand in the near future (PI_1)	[6]	5-pt

Note. Representative items are presented due to space limitations, while all construct indicators were included in the measurement model and evaluated through outer loadings. SMC = Social Media Content.

Common method bias (CMB) was assessed using both procedural and statistical remedies. Procedurally, two questionnaire versions were prepared by reordering construct blocks, and attention-check items were embedded to reduce response pattern bias. Statistically, Harman's single-factor test was conducted to examine whether a single factor dominated the covariance among measurement items. The first unrotated factor accounted for 28.6% of the total variance, which is below the commonly used 50% threshold, suggesting that common method bias was unlikely to pose a serious threat to the validity of the findings [14]. Nevertheless, because Harman's test is only a diagnostic procedure, the results were interpreted cautiously and complemented by the procedural safeguards applied during questionnaire design.

2.4 Analytical Procedure

PLS-SEM was estimated in SmartPLS 4 using the two-step protocol of Hair et al.: convergent validity (outer loadings ≥ 0.70, AVE ≥ 0.50), internal consistency (CR ≥ 0.70), discriminant validity (HTMT < 0.85) [15], and structural paths via 5,000-resample bootstrapping. PLSpredict computed Q^2_predict for out-of-sample relevance [8]. PLS-SEM latent scores were then standardized (mean = 0, SD = 1) and split into training (80%, n = 400) and test (20%, n = 100) sets using stratified random sampling on PI quartiles. Three ML models were estimated: Linear Regression as a parametric baseline [16], Random Forest as a bagging benchmark, and XGBoost as the primary model for its regularization properties and TreeSHAP compatibility [9]. XGBoost hyperparameters were tuned via exhaustive grid search with 10-fold cross-validation (Table 3). SHAP values were computed using TreeExplainer [17], which decomposes each prediction into additive feature contributions satisfying efficiency, symmetry, and dummy compliance [10].

Table 3. XGBoost Hyperparameter Grid Search and Optimal Configuration

Parameter	Grid Values Tested	Optimal
n_estimators	100, 200, 300	200
max_depth	3, 5, 7	3
learning_rate	0.01, 0.05, 0.1	0.05
subsample	0.7, 0.8, 1.0	0.8
colsample_bytree	0.7, 0.8, 1.0	0.8
reg_alpha	0, 0.1, 0.5	0.1
reg_lambda	1, 1.5, 2	1.5

Note. Model selection based on minimizing mean CV-RMSE across 10 folds on training set (n = 400).

3. Results and Discussion

3.1 Measurement Model

All indicator loadings exceeded 0.70 (range: 0.830–0.865 for SMC; 0.805–0.835 for BI; 0.838–0.884 for PI), AVE values ranged from 0.672 to 0.746, and CR from 0.922 to 0.942 (Table 4) confirming convergent validity and internal consistency across all constructs [12], [18]. Among SMC indicators, IG_CRE and IG_ENT carried the highest loadings (0.865 and 0.855), consistent with Tafesse and Wien [3] who found that creative and entertainment dimensions drive stronger brand engagement on visual platforms than informational content alone.

Table 4. Outer Loadings, AVE, and Composite Reliability

Construct	Indicator	Loading	AVE	CR
Social Media Content	IG_CRE	0.865	0.711	0.925
	IG_ENT	0.855		
	IG_INF	0.851		
	IG_INT	0.832		
	IG_VIS	0.830		
Brand Image	BI_1–BI_8	0.805–0.835	0.672	0.942
Purchase Intention	PI_1–PI_4	0.838–0.884	0.746	0.922

Note. Loadings > 0.70; AVE > 0.50; CR > 0.70. Thresholds per Hair et al. [19] and Meshage et al. [18]

Discriminant validity was confirmed via HTMT (Table 5); the highest ratio was 0.720 (SMC–BI), well below the 0.85 threshold [15], establishing that the three constructs are empirically distinguishable despite their conceptual proximity.

Table 5. Discriminant Validity HTMT Matrix

Construct	SMC	Brand Image	Purch. Intention
Social Media Content		0.720	0.597
Brand Image	0.720		0.624
Purchase Intention	0.597	0.624	

Note. All HTMT values < 0.85 [15].

3.2 Structural Model and Hypothesis Testing

Table 6 reports R^2 values and bootstrap path coefficients. H1 is supported (SMC → BI: $\beta = 0.581$, $t = 19.67$, $p < .001$): social media content accounts for 33.8% of brand image variance, a moderate level consistent with comparable digital marketing studies [20]. The notably high t-statistic reflects precision across 5,000 bootstrap samples and confirms that the SMC–BI relationship is not resampling-sensitive. H2 is also supported (BI → PI: $\beta = 0.511$, $t = 14.40$, $p < .001$): consumers holding more positive brand associations are more likely to express purchase intent consistent with the theoretical reasoning of Keller and Swaminathan [5] and online purchasing evidence from [6]. The R^2 for PI (0.261) falls in the low-to-moderate range, which is expected: purchase intention is shaped by factors beyond brand image and content, including price sensitivity, social norms, and platform trust. Comparable SEM studies routinely report PI R^2 between 0.20 and 0.35 [6], [20], placing the current value within an acceptable range.

Table 6. Structural Path Coefficients and R^2 (5,000 Bootstrap Resamples)

Path / Construct	β	t-stat	p	Decision
H1: SMC → Brand Image	0.581	19.67	< .001	Supported
H2: Brand Image → PI	0.511	14.40	< .001	Supported
R^2: Brand Image (0.338)				Moderate
R^2: Purchase Intention (0.261)				Low–Mod.

Note. Shaded rows = R^2 summaries. β = standardized coefficient; p-values from two-tailed bootstrap tests.

H3 is supported: brand image significantly mediates the relationship between social media content and purchase intention (indirect $\beta = 0.297$, 95% BC-CI: [0.241, 0.356], Table 7). The confidence interval excludes zero, indicating that the indirect effect is statistically significant and stable across bootstrap resamples. This finding suggests that improvements in social media content contribute to purchase intention primarily by strengthening brand image, making brand perception an important psychological mechanism through which digital content influences consumer decision-making [5], [21].

Table 7. Mediation Analysis Indirect Effect (5,000 Bootstrap Resamples)

Path	β	95% BC-CI	Mediation
SMC → Brand Image → PI	0.297	[0.241, 0.356]	Significant mediation

Note. BC-CI = bias-corrected bootstrap confidence interval. The confidence interval excludes zero, confirming a significant indirect effect.

PLSpredict Q^2_predict values were positive for both constructs (BI: 0.173–0.237, moderate; PI: 0.137–0.143, low), confirming that the SEM model outperforms a naïve benchmark [8]. The lower Q^2 for PI reflects its distal nature as a behavioral outcome shaped by many factors outside the model and provided the empirical rationale for extending the analysis with machine learning.

3.3 Machine Learning Performance and Cross-Validation

Table 8 presents training and test performance for all three models. Random Forest's dramatic overfitting (train $R^2 = 0.827$, test $R^2 = -0.045$) is consistent with its behavior in sparse feature spaces: with only two input features, its high-variance trees memorize noise rather than capturing signal, and the absence of regularization prevents generalization [16]. The near-identical test performance of Linear Regression ($R^2 = 0.288$) and XGBoost ($R^2 = 0.277$) warrants explanation. With only two construct-level features themselves composite-weighted linear combinations of indicator items most non-linear interaction structure is already suppressed before it enters the model. Purchase intention's restricted five-point scale further limits the variance ceiling. In this configuration, XGBoost's value lies not in accuracy gains over a linear baseline but in its compatibility with TreeSHAP interpretability [10], [17].

Table 8. ML Model Performance Training vs. Test Set

Model / Construct	Train RMSE	Test RMSE	Train R^2	Test R^2
Linear Regression	1.028	0.986	0.296	0.288
Random Forest	0.510	1.194	0.827	−0.045
XGBoost (tuned)	0.994	0.994	0.343	0.277

Note. Shaded row = selected primary model. Test set n = 100. Metrics computed on standardized PI scores.

Ten-fold stratified cross-validation (Table 9) confirmed these patterns. Linear Regression (CV $R^2 = 0.273$) and XGBoost (CV $R^2 = 0.267$) both showed stable fold-to-fold performance (SD ≈ 0.09), ruling out favorable data splits as an explanation. Random Forest's CV R^2 collapsed to 0.014 with higher variance across folds, reinforcing the overfitting diagnosis and eliminating it from further consideration [16].

Table 9. 10-Fold Cross-Validation Results (Training Set, n = 400)

Model	CV RMSE (mean±SD)	CV R^2	Stability
Linear Regression	1.024 ± 0.091	0.273	Stable
Random Forest	1.192 ± 0.143	0.014	Unstable
XGBoost (tuned)	1.030 ± 0.094	0.267	Stable

Note. 'Stable' = fold SD < 0.10; 'Unstable' = fold SD ≥ 0.10.

3.4 SHAP Analysis and Cross-Method Comparison

Brand image emerged as the stronger predictive feature, with a mean |SHAP| value of 0.302, followed by social media content with a mean |SHAP| value of 0.268. Because mean |SHAP| values represent the magnitude of predictive contribution rather than directional effects, these values should not be interpreted as equivalent to SEM path coefficients. Nevertheless, the SHAP pattern was consistent with the positive signs observed in the SEM structural model, indicating that the predictive contribution of both constructs aligns with the theoretical relationships tested in PLS-SEM. Table 10 places these SHAP values alongside SEM path coefficients, enabling the core cross-method comparison of this study.

Table 10. Cross-Method Comparison SEM Path Coefficients vs. SHAP Importance

Variable	SEM β	Mean |SHAP|	SHAP Rank
SMC → Brand Image	0.581	0.268	2nd
Brand Image → PI	0.511	0.302	1st

Note. SEM β and SHAP values are not directly comparable in magnitude they measure different quantities but their rank ordering can be meaningfully compared. SMC = Social Media Content; PI = Purchase Intention.

Construct-level SHAP was intentionally used because the primary objective of this study was to compare theory-derived latent constructs across explanatory and predictive paradigms, rather than to optimize item-level prediction. This design choice allows SHAP values to be interpreted at the same conceptual level as SEM path coefficients, thereby making the cross-method comparison more theoretically defensible. Nevertheless, item-level SHAP remains a promising extension for future research.

Three observations emerge. First, both methods converge on which constructs matter, providing mutual validation: the SEM result is not an artifact of model constraints, and the XGBoost finding is not disconnected from theory. Second, the rank ordering inverts between paradigms: SEM assigns the larger coefficient to SMC ($\beta = 0.581$) while SHAP ranks BI as the stronger predictor of PI (0.302 vs. 0.268). This inversion is theoretically coherent rather than contradictory it reflects the distinction between causal salience, namely how much one construct drives another as captured by SEM, and predictive prominence, namely how much a construct contributes to forecasting a downstream outcome as captured by SHAP. Brand image acts as an amplifier: social media content initiates it, but brand image is what ultimately translates into purchase intention [5], [10]. Third, the compression in SHAP magnitude differences (0.034 gap vs. 0.070 gap in SEM) reflects even credit distribution across two features and the homogenized input signal produced by latent scores an inherent property of the hybrid design that reinforces why direct numerical comparison of SEM and SHAP values is inappropriate [10], [17].

For digital marketing practitioners, the combined findings translate into two actionable priorities. First, content investment should emphasize creative and entertainment dimensions which carried the highest indicator loadings over purely informational formats, as these dimensions most effectively shift brand perception. Second, and more strategically, brand image maintenance must be treated as an ongoing program rather than a campaign-by-campaign output: the SHAP evidence shows it is the in-memory brand associations that most strongly predict whether consumers buy, meaning that brands allowing image erosion between campaigns sacrifice the very mechanism through which content investments convert into revenue.

4. Conclusions

This study integrated PLS-SEM, XGBoost, and SHAP to examine how social media content shapes purchase intention through brand image mediation and whether the structural relationships identified by the confirmatory model withstand predictive scrutiny. All three hypotheses were supported: SMC → BI ($\beta = 0.581$), BI → PI ($\beta = 0.511$), and brand image as a significant mediator (indirect $\beta = 0.297$, 95% BC-CI: [0.241, 0.356]). In the predictive stage, XGBoost and Linear Regression generalized stably (test $R^2 \approx 0.28$) while Random Forest overfit severely (test $R^2 = -0.045$). SHAP analysis ranked brand image as the stronger predictive feature (0.302 > 0.268), with the SHAP pattern remaining consistent with the positive signs observed in the SEM structural model.

The study's primary theoretical contribution is the articulation of *causal salience versus predictive prominence* as distinct and non-redundant properties of constructs in behavioral models. SEM identifies social media content as the stronger causal driver (higher β toward brand image), while SHAP identifies brand image as the stronger predictive feature for the downstream outcome. A variable can be pivotal in moving another construct yet carry less direct predictive weight for a distal behavioral outcome than the construct it shapes a conceptual distinction not previously articulated in hybrid SEM–ML marketing research. Methodologically, the study demonstrates that using PLS-SEM latent scores as ML inputs produces psychometrically validated features that make SHAP outputs more interpretable and the cross-method comparison more defensible [12], [22]; it also shows that in low-dimensional construct-level data, XGBoost's advantage over linear models lies in interpretability rather than accuracy [9], [10], [23].

Several limitations bound these findings. The purposive sample from Indonesian Instagram users limits generalizability across cultural and platform contexts. The cross-sectional, single-source design despite a Harman test first-factor loading of 28.6% cannot fully exclude common method variance, particularly between the two attitudinal constructs. Predictive performance is modest (test $R^2 \approx 0.28$), partly because construct-level latent scores compress the item-level variance that would enable non-linear models to outperform linear baselines, and partly because purchase intention is inherently influenced by factors price, social norms, platform trust outside the current model. Finally, SHAP was applied at the construct level only; item-level decomposition would reveal which specific content dimensions and brand image facets carry the greatest predictive weight, and this extension is the most immediate direction for future work.

Future research should retain item-level inputs for the ML stage to exploit XGBoost's non-linear capacity and enable finer-grained SHAP interpretation. Incorporating trust [6], perceived value, and platform-specific algorithmic exposure as additional constructs would address the model's explanatory ceiling. A longitudinal design tracking respondents across multiple Instagram exposure waves would allow the temporal unfolding of the content–image–intention chain to be tested directly rather than inferred from cross-sectional associations. Cross-platform replication across TikTok, YouTube, and X, and extension to B2B marketing or technology adoption domains, would establish the external validity of the hybrid explainable framework.

In closing, the divide between theory-testing and prediction in consumer behavior research is a methodological convention, not an empirical necessity. When SEM and machine learning are designed to work together with construct validation preceding ML feature construction and SHAP bridging the outputs the combined framework surfaces findings that neither method could produce independently. The convergence between SEM structural paths and SHAP predictive rankings observed here demonstrates that theory-derived constructs retain explanatory relevance under predictive scrutiny, and that theoretically grounded features produce more interpretable ML models than raw survey items would. These observations strengthen the case for hybrid explainable approaches as a standard analytical toolkit in empirical marketing research.

References

[1] Y. K. Dwivedi *et al.*, “Opinion Paper: ‘So what if ChatGPT wrote it?’ Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy,” *International Journal of Information Management*, vol. 71, p. 102642, Aug. 2023, doi: 10.1016/j.ijinfomgt.2023.102642.

[2] M. Yadav and Z. Rahman, "The influence of social media marketing activities on customer loyalty," *Benchmarking: An International Journal*, vol. 25, no. 9, pp. 3882–3905, Nov. 2018, doi: 10.1108/BIJ-05-2017-0092.

[3] W. Tafesse and A. Wien, "Implementing social media marketing strategically: an empirical assessment," *Journal of Marketing Management*, vol. 34, no. 9–10, pp. 732–749, Jun. 2018, doi: 10.1080/0267257X.2018.1482365.

[4] D. Vrontis, A. Makrides, M. Christofi, and A. Thrassou, "Social media influencer marketing: A systematic review, integrative framework and future research agenda," *Int J Consumer Studies*, vol. 45, no. 4, pp. 617–644, Jul. 2021, doi: 10.1111/ijcs.12647.

[5] K. L. Keller and V. Swaminathan, *Strategic brand management: building, measuring, and managing brand equity*, Fifth edition, Global edition. Harlow, England London New York: Pearson, 2020.

[6] C. Hsu, J. Chuan-Chuan Lin, and H. Chiang, "The effects of blogger recommendations on customers' online shopping intentions," *Internet Research*, vol. 23, no. 1, pp. 69–88, Jan. 2013, doi: 10.1108/10662241311295782.

[7] N. Wathanakom, "Structural Equation Modelling in Marketing: A Systematic Review of Methods and Models," *ecrm*, vol. 24, no. 1, pp. 139–145, Jun. 2025, doi: 10.34190/ecrm.24.1.3635.

[8] G. Shmueli *et al.*, "Predictive model assessment in PLS-SEM: guidelines for using PLSpredict," *EJM*, vol. 53, no. 11, pp. 2322–2347, Nov. 2019, doi: 10.1108/EJM-02-2019-0189.

[9] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[10] Christoph Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (3rd ed.)*, 3rd ed. https://christophm.github.io/interpretable-ml-book/, 2025. Accessed: Apr. 08, 2025. [Online]. Available: https://christophm.github.io/interpretable-ml-book/

[11] Hudhaifa Ayad, "HYBRID STRUCTURAL EQUATION MODELING AND MACHINE LEARNING APPROACHES FOR PREDICTING ACADEMIC OUTCOMES: A LITERATURE REVIEW IN IRAQI HIGHER EDUCATION E-LEARNING," *ijam*, vol. 38, no. 9s, pp. 139–161, Oct. 2025, doi: 10.12732/ijam.v38i9s.770.

[12] V. Ferdinand and S. F. Wijaya, "Integrating SmartPLS-Based Structural Modeling for Predictive Machine Learning in Retail Customer Behavior," in *2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Jakarta, Indonesia: IEEE, Dec. 2025, pp. 1879–1885. doi: 10.1109/ICIMCIS68501.2025.11327115.

[13] J. Li, Y. Horiguchi, and T. Sawaragi, "Data Dimensionality Reduction by Introducing Structural Equation Modeling to Machine Learning Problems," in *2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, Chiang Mai, Thailand: IEEE, Sep. 2020, pp. 826–831. doi: 10.23919/SICE48898.2020.9240323.

[14] P. M. Podsakoff, S. B. MacKenzie, J.-Y. Lee, and N. P. Podsakoff, "Common method biases in behavioral research: A critical review of the literature and recommended remedies.," *Journal of Applied Psychology*, vol. 88, no. 5, pp. 879–903, 2003, doi: 10.1037/0021-9010.88.5.879.

[15] E. Roemer, F. Schuberth, and J. Henseler, "HTMT2–an improved criterion for assessing discriminant validity in structural equation modeling," *IMDS*, vol. 121, no. 12, pp. 2637–2650, Nov. 2021, doi: 10.1108/IMDS-02-2021-0082.

[16] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*. in Springer Texts in Statistics. New York, NY: Springer US, 2021. doi: 10.1007/978-1-0716-1418-1.

[17] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nat Mach Intell*, vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9.

[18] J. Meshage, J. Alias, M. M. Awang, and W. Yang, "Development and validation of a scale for measuring social environmental support, survival skills, and subjective well-being among Chinese ethnic minority university students," *ijirss*, vol. 8, no. 6, pp. 2538–2548, Sep. 2025, doi: 10.53894/ijirss.v8i6.10137.

[19] J. F. Hair, Ed., *A primer on partial least squares structural equation modeling (PLS-SEM)*, Second edition. Los Angeles: Sage, 2017.

[20] O. Vania, "THE INFLUENCE OF SOCIAL MEDIA CONTENT ON BRAND IMAGE, BRAND AWARENESS, AND PURCHASE DECISIONS (Case Study on Brand Berrybenka)," *METADATA*, vol. 7, no. 3, pp. 1–10, Sep. 2025, doi: 10.47652/metadata.v7i3.600.

[21] Razin Auliaur Al-Asyraf, Siti Mujanah, and Achmad Yanu Alifianto, "Examining the Influence of Product Quality on Repurchase Intention: The Mediating Roles of Brand Image and Word of Mouth in the Context of Weber Products," *IJECM*, vol. 2, no. 4, pp. 242–254, Oct. 2025, doi: 10.62951/ijecm.v2i4.1092.

[22] C. Magasi and A. M. Nyamwesa, "Reimagining Consumer Analytics: Predictive and Real-Time Insights Through Dynamic Structural Equation Modeling," *IJBE*, vol. 11, no. 3, p. 734, Sep. 2025, doi: 10.17358/ijbe.11.3.734.

[23] S. Bartoloni and C. Ancillai, "Twenty years of social media marketing: A systematic review, integrative framework, and future research agenda," *Int J Management Reviews*, vol. 26, no. 3, pp. 435–457, Jul. 2024, doi: 10.1111/ijmr.12360.