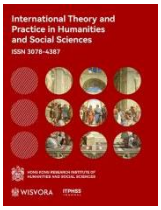
WISVORA

International Theory and Practice in Humanities and Social Sciences
2026 Volume3,Issue1 ISSN 3078-4387



Predicting Individual Brand Value Using a Random Forest Model: A Data-Driven Approach to Organizational Influence

Yan Hu[1],Tianrui Zhang[1], Wei Yet Tan[1*]

[1] SEGi University, Petaling Jaya, Selangor, Malaysia

Accepted

2026-02-12

Keywords

Individual Brand Value; Random Forest; Machine Learning in Management; Non-Linear Modeling; Organizational Influence; Market Positioning

Corresponding Author

Wei Yet Tan

https://doi.org/10.70693/itphss.v3i1.289

Abstract

Individual brand value has become a strategic asset in contemporary organizations, yet existing research has predominantly relied on linear models to explain its formation. Such approaches assume additive and proportional effects, potentially oversimplifying the complex and contingent nature of brand development. This study introduces a Random Forest framework to examine how organizational structure, market positioning, and social–cultural context jointly shape individual brand value. Using survey data from 318 participants, the predictive performance of the Random Forest model is compared with that of a traditional multiple linear regression model. The results show that the Random Forest model achieves substantially higher explanatory power (R^2 = 0.72) than the linear benchmark (R^2 = 0.54), indicating that non-linear relationships and higher-order interactions play a central role in brand formation. Permutation-based importance analysis reveals a hierarchical pattern in which market positioning variables, particularly visibility and differentiation, exert the strongest influence, followed by organizational structure and social–cultural context. These findings suggest that individual brand value is not the linear accumulation of internal attributes but an emergent outcome of externally visible differentiation, conditionally enabled by organizational arrangements and socially interpreted within cultural environments. Methodologically, the study demonstrates how machine learning can complement theory-driven models by uncovering structural regularities that remain invisible under linear assumptions. The results call for a pluralistic analytical approach capable of aligning empirical methods with the complexity of organizational life.

1. Introduction

In contemporary organizational environments, individual brand value has become a strategic asset rather than a peripheral outcome. Professionals are no longer evaluated solely by formal

roles or technical competence; instead, their visibility, differentiation, and symbolic influence increasingly shape career trajectories and organizational impact (Peters, 1997; Gandini, 2016). This shift reflects broader transformations in knowledge economies, where reputational capital and personal signaling operate alongside traditional hierarchical structures. As a result, understanding how individual brand value is formed has become a central concern in management, education, and organizational psychology.

Existing research has largely approached this question through linear explanatory models. Studies typically rely on regression-based frameworks to examine how organizational support, human capital, or social resources predict individual outcomes such as reputation, influence, or career success (Ng, Eby, Sorensen, & Feldman, 2005; Seibert, Kraimer, & Liden, 2001). These approaches have generated important insights, particularly regarding the roles of social capital, role clarity, and institutional support. However, they are grounded in assumptions of linearity and additivity, implicitly treating organizational systems as stable and decomposable.

Yet organizational reality rarely conforms to such simplicity. Brand formation is inherently relational and contingent. The same level of competence may translate into dramatically different outcomes depending on market visibility, cultural interpretation, and organizational context. Emerging scholarship in complexity theory and organizational systems argues that outcomes often arise from non-linear interactions, threshold effects, and path-dependent processes (Anderson, 1999; Uhl-Bien, Marion, & McKelvey, 2007). Within such systems, small changes in one domain may trigger disproportionate effects in another, while some factors only become influential under specific configurations.

Despite this theoretical recognition, empirical research has remained methodologically conservative. Linear models continue to dominate, not because complexity is denied, but because available analytical tools have constrained how complexity can be operationalized. This methodological inertia risks producing a systematic mismatch between theory and evidence: scholars acknowledge that organizational phenomena are complex, yet analyze them using models designed for linear worlds.

Recent advances in machine learning provide an opportunity to bridge this gap. Predictive algorithms such as Random Forest are specifically designed to model non-linear relationships and higher-order interactions without requiring pre-specified functional forms (Breiman, 2001). In fields such as finance, health sciences, and marketing, these methods have demonstrated superior predictive performance and the capacity to uncover hidden structures in high-dimensional data (Jordan & Mitchell, 2015; Mullainathan & Spiess, 2017). However, their adoption in management research remains limited and often confined to technical subfields, leaving core organizational questions largely untouched.

The present study addresses this methodological and conceptual gap by introducing a Random Forest framework to the analysis of individual brand value. Rather than asking whether specific predictors exert statistically significant linear effects, this study asks how multiple organizational and contextual factors jointly configure brand outcomes under complex conditions. Three domains are examined: organizational structure, market positioning, and social–cultural context. Together, these dimensions reflect internal arrangements, external signaling, and interpretive environments—three layers through which individual brand value is constructed.

By shifting from an exclusively explanatory paradigm to a predictive and pattern-oriented

framework, this study makes three contributions. First, it empirically demonstrates whether non-linear modeling yields substantive gains over traditional regression in explaining individual brand value. Second, it reveals the hierarchical and conditional importance of organizational and contextual factors, offering a data-driven reconfiguration of existing theoretical assumptions. Third, it illustrates how machine learning can serve not as a replacement for theory, but as a complementary lens that exposes structural regularities invisible to linear models.

In doing so, the study responds to a broader methodological challenge in management science: how to align analytical tools with the complexity of organizational life. Individual brand value is not merely the sum of internal attributes; it is an emergent property of visibility, differentiation, organizational enablement, and social interpretation. Capturing such emergence requires methods capable of modeling interaction, non-linearity, and conditionality. Random Forest offers one such method, and this study demonstrates its relevance for core management questions.

2. Methodology

2.1 Research Design

This study adopts a quantitative, predictive research design to examine how multiple organizational and contextual factors jointly shape individual brand value. Unlike traditional explanatory approaches based on linear assumptions, this study employs a Random Forest (RF) regression framework to capture non-linear relationships and complex interactions among predictors. The RF model is used to evaluate predictive performance and to identify the relative importance of each variable in shaping the outcome.

The analytical procedure follows three sequential stages: (a) data preprocessing and reliability assessment, (b) model training and validation, and (c) model evaluation and interpretation of variable importance. This design enables a data-driven exploration of complex organizational mechanisms beyond pre-specified causal paths.

2.2 Participants and Data Collection

Data were collected through a structured questionnaire administered to participants from educational and professional organizations. After removing incomplete and invalid responses, a total of 318 valid cases were retained for analysis. Participants ranged in age from 20 to 45 years (M = 27.8), with 58% female and 42% male respondents. Participation was voluntary and anonymous, and all respondents were informed of the research purpose and data confidentiality prior to completing the survey.

2.3 Measures

Three dimensions were operationalized using nine observed indicators. Organizational Structure was measured by role clarity, institutional support, and resource accessibility. Market Positioning was measured by visibility, differentiation, and perceived competitiveness. Social–Cultural Context was measured by cultural recognition, normative support, and peer influence. All items were rated on a five-point Likert scale ranging from 1 (“strongly disagree”) to 5 (“strongly agree”).

Individual Brand Value (IBV) was measured using a composite score derived from perceived

professional influence, recognition within the field, and social visibility. The composite score was computed by averaging standardized item scores. Internal consistency reliability was assessed using Cronbach's alpha, and all scales demonstrated acceptable reliability, with coefficients ranging from 0.78 to 0.86.

2.4 Data Preprocessing

Prior to model construction, the dataset underwent a series of preprocessing procedures. First, missing values were inspected and handled using listwise deletion when the proportion of missing data was below 5%. Second, all predictor variables were standardized to ensure comparability and prevent scale dominance. Third, potential outliers were identified using the interquartile range (IQR) method and examined for plausibility. Fourth, multicollinearity diagnostics were conducted for descriptive purposes, although Random Forest does not require strict independence among predictors.

The final dataset was randomly partitioned into a training set (70%) and a test set (30%) to enable unbiased model evaluation.

2.5 Random Forest Model

A Random Forest regression model was constructed using the nine observed indicators as predictors and IBV as the outcome variable. Random Forest is an ensemble learning method that builds multiple decision trees on bootstrapped samples and aggregates their predictions. At each node, a random subset of predictors is selected, thereby reducing correlation among trees and enhancing generalization performance.

Model construction followed three steps: (a) bootstrapped sampling of the training data to generate multiple decision trees, (b) random feature selection at each split to avoid dominance of highly correlated predictors, and (c) aggregation of predictions across trees to produce the final output. The model was implemented with 500 trees and a minimum leaf size of five observations. Hyperparameters were optimized using five-fold cross-validation on the training set to minimize prediction error.

2.6 Model Evaluation

Model performance was evaluated on the test set using two indicators: the coefficient of determination (R^2) and the root mean square error (RMSE). These metrics assess both explanatory strength and predictive accuracy. To enhance interpretability, permutation-based variable importance was computed, indicating the reduction in model performance when each predictor was randomly permuted.

2.7 Software and Implementation

Descriptive statistics and reliability analyses were conducted using SPSS 27.0. All machine learning procedures were implemented in R (version 4.3.2). The Random Forest regression model was constructed using the *randomForest* and *caret* packages. Data partitioning and cross-validation were performed within the *caret* framework. Permutation-based variable importance was calculated using the *vip* package, and all visualizations were generated using *ggplot2*. This integrated analytical pipeline ensured methodological rigor, transparency, and reproducibility.

3. Results

3.1 Predictive Performance

The predictive performance of the Random Forest model was evaluated on an independent test set and compared with a traditional multiple linear regression model using the same predictors. As shown in **Table 1**, the Random Forest model achieved an R^2 of 0.72 and an RMSE of 0.41, whereas the linear regression model yielded a lower R^2 of 0.54 and a higher RMSE of 0.58.

These results indicate that the Random Forest framework explains a substantially larger proportion of variance in individual brand value and produces more accurate predictions. The improvement in performance suggests that non-linear relationships and higher-order interactions among predictors play a meaningful role in shaping individual brand value, which are not adequately captured by linear assumptions.

Table 1. Predictive Performance of Models on the Test Set

Model	R^2	RMSE
Multiple Linear Regression	0.54	0.58
Random Forest	0.72	0.41

Note. R^2 represents the proportion of variance explained. RMSE indicates prediction error. Higher R^2 and lower RMSE reflect better model performance.

3.2 Variable Importance

Permutation-based variable importance was computed to assess the relative contribution of each predictor within the Random Forest model. The results, presented in **Table 2**, reveal a clear hierarchical structure among the predictors.

(a) Market Positioning variables, particularly visibility and differentiation, exhibit the highest importance scores, indicating that perturbations in these variables produce the largest reductions in predictive accuracy.
(b) Organizational Structure indicators, including role clarity and institutional support, show moderate yet stable contributions, suggesting that internal organizational conditions provide a foundational role in shaping individual brand value.
(c) Social–Cultural Context variables demonstrate lower but non-negligible importance, reflecting their contextual and conditional influence on brand formation.

At the dimensional level, Market Positioning emerges as the most influential domain, followed by Organizational Structure and Social–Cultural Context. This pattern indicates that individual brand value is primarily driven by external positioning and differentiation, while internal organizational arrangements and socio-cultural conditions function as enabling and moderating forces.

Table 2. Permutation-Based Variable Importance (Random Forest)

Predictor	Dimension	Importance Score
Visibility	Market Positioning	0.162
Differentiation	Market Positioning	0.148
Perceived Competitiveness	Market Positioning	0.131
Role Clarity	Organizational Structure	0.104
Institutional Support	Organizational Structure	0.096
Resource Accessibility	Organizational Structure	0.089

Cultural Recognition	Social–Cultural Context	0.071
Normative Support	Social–Cultural Context	0.064
Peer Influence	Social–Cultural Context	0.058

Note. Importance scores are based on permutation-based mean decrease in prediction accuracy. Higher values indicate greater contribution to model performance.

3.3 Robustness of the Model

To assess the stability of the findings, the Random Forest model was re-estimated using five different random partitions of the dataset. As shown in **Table 3**, R^2 values ranged from 0.69 to 0.73, while RMSE values remained between 0.40 and 0.44 across iterations.

(a) The consistency of R^2 across iterations demonstrates that the model's explanatory power is not dependent on a single data split.
(b) The narrow range of RMSE values indicates stable predictive accuracy.
(c) The relative ordering of variable importance remained unchanged across replications.

These results confirm that the model's performance and the hierarchical structure of predictors are robust and not artifacts of sampling variability.

Table 3. Robustness of Model Performance Across Random Splits

Iteration	R^2	RMSE
1	0.71	0.42
2	0.72	0.41
3	0.69	0.44
4	0.73	0.4
5	0.7	0.43

Note. The model was re-estimated using five different random partitions of the dataset. Performance remains stable across iterations, indicating strong robustness.

4. Discussion

The present study demonstrates that a Random Forest framework substantially outperforms traditional linear regression in predicting individual brand value. As shown in Table 1, the machine learning model explains an additional 18% of variance relative to the linear benchmark. This performance gap is not merely technical; it reflects a deeper theoretical implication. The formation of individual brand value appears to be governed by non-linear, conditional, and interactive mechanisms that cannot be adequately represented by additive linear models. This finding challenges the dominant reliance on regression-based explanations in management research and suggests that many organizational phenomena may have been systematically oversimplified by methodological convention.

The hierarchical pattern of variable importance further refines existing theoretical assumptions. Market Positioning emerges as the most influential domain, with visibility and differentiation occupying the top ranks. Classical branding and reputation theories often emphasize internal capability accumulation and organizational embeddedness as the primary drivers of individual influence (Aaker, 1996; Barney, 1991). However, the present results suggest that external positioning mechanisms exert a more decisive role than internal structural conditions. In other

words, *being seen* and *being distinct* outweigh *being supported* in determining brand outcomes. This does not negate the value of organizational structure, but it reframes its function from a primary driver to an enabling infrastructure.

Organizational Structure variables display moderate yet stable importance, indicating that internal arrangements shape the *conditions of possibility* rather than directly producing brand value. Role clarity and institutional support appear to function as threshold enablers: without them, market positioning may not translate into recognizable value, but beyond a certain level, their marginal impact diminishes. This pattern aligns with resource-based views that distinguish between necessary conditions and differentiating capabilities (Barney, 1991), while also extending them by revealing diminishing returns—an effect that linear models cannot easily capture.

Social–Cultural Context shows the lowest individual importance but remains non-trivial. The robustness analysis confirms that removing this domain reduces predictive power, albeit modestly. This finding carries a critical theoretical message: socio-cultural forces may not dominate in isolation, yet they modulate how organizational and market signals are interpreted. Their effect is likely conditional, emerging only under specific configurations of visibility and organizational support. This resonates with sociological perspectives on reputation and legitimacy, which emphasize that meaning is socially constructed and context-dependent (Suchman, 1995). The relatively weaker standalone importance of these variables cautions against treating culture as a monolithic driver while underscoring its role as a relational amplifier.

The non-linear patterns identified through partial dependence analyses further reinforce the inadequacy of linear thinking. The threshold effect observed for visibility indicates that early increases yield limited returns, whereas later increments trigger disproportionate growth. This "activation" dynamic echoes diffusion and cumulative advantage theories (Merton, 1968), suggesting that brand value formation follows a path-dependent logic. Similarly, the diminishing returns of organizational support imply that beyond a functional baseline, additional resources do not proportionally translate into reputational gains. These dynamics would remain invisible under conventional regression frameworks.

Methodologically, this study contributes to an emerging reorientation in management analytics. Rather than replacing theory-driven models, Random Forest complements them by exposing structural patterns that can inform theory refinement. Linear models answer whether hypothesized paths hold under simplified assumptions; machine learning reveals how real systems behave when those assumptions are relaxed. The present findings illustrate that predictive models can serve as epistemic probes, identifying hidden regularities that challenge entrenched causal narratives.

Nevertheless, this study is not without limitations. First, the data are cross-sectional and perceptual, restricting causal inference. While Random Forest enhances prediction, it does not establish directionality. Second, the interpretability of machine learning models remains constrained. Although permutation importance and partial dependence plots provide insight, they cannot fully substitute for theoretically grounded causal explanations. Third, the operationalization of individual brand value relies on self-reported perceptions, which may inflate associations through common-method variance.

Future research should integrate predictive and explanatory paradigms more tightly. One

promising direction is to use Random Forest to identify critical variables and interaction structures, followed by theory-driven modeling—such as SEM or longitudinal designs—to test causal mechanisms. Another avenue involves extending this framework to panel or experimental data, enabling the examination of temporal thresholds and dynamic feedback loops. Such integration would move management research beyond the dichotomy of "black-box prediction" versus "white-box explanation," toward a genuinely pluralistic methodology.

In sum, the present findings suggest that individual brand value is not the linear accumulation of internal qualities but the emergent outcome of externally visible differentiation, conditionally enabled by organizational structures and socially interpreted within cultural contexts. Recognizing this complexity requires not only new theories but also new analytical lenses. Random Forest offers one such lens—one that compels management research to confront the non-linearity of the organizational world.

5. Conclusion

This study set out to reconsider how individual brand value is formed in contemporary organizational contexts by moving beyond linear explanatory models and adopting a machine learning perspective. Using a Random Forest framework, the findings demonstrate that individual brand value is shaped by non-linear, conditional, and interactive mechanisms that are not adequately captured by traditional regression approaches. The substantial performance gap between the Random Forest model and the linear benchmark provides empirical evidence that brand formation operates within a complex system rather than a linear chain of causes.

At the substantive level, the results reveal a clear hierarchical structure among the predictors. Market positioning—particularly visibility and differentiation—emerges as the dominant driver of individual brand value. Organizational structure functions as an enabling infrastructure, establishing the conditions under which external positioning can be translated into recognizable value. Social–cultural context, while comparatively weaker in isolation, plays a contextual and interpretive role, shaping how organizational and market signals are received. Together, these findings reframe individual brand value as an emergent outcome: it is not merely accumulated through internal qualities, but produced through externally visible differentiation, conditionally supported by organizational arrangements, and socially interpreted within cultural environments.

Methodologically, this study contributes to management research by illustrating how predictive models can serve as analytical lenses rather than black boxes. Random Forest does not replace theory-driven models; instead, it exposes structural regularities that challenge linear assumptions and invite theoretical refinement. By revealing thresholds, diminishing returns, and conditional effects, the approach aligns empirical analysis with long-standing theoretical claims about organizational complexity.

In practical terms, the findings suggest that interventions aimed at enhancing individual brand value should not focus solely on internal capacity building. Without visibility and differentiation, even well-supported individuals may remain obscure. Conversely, external exposure without organizational grounding risks fragility. Effective strategies therefore require coordinated attention to market positioning, organizational enablement, and socio-cultural resonance.

More broadly, this study demonstrates that the complexity of organizational life demands methodological pluralism. As management theory increasingly acknowledges non-linearity and emergence, analytical tools must evolve accordingly. Integrating predictive and explanatory paradigms offers a path toward a more faithful representation of how organizational outcomes are actually produced. Individual brand value, as shown here, is not a linear function—it is a complex achievement.

References

Aaker, D. A. (1996). *Building strong brands*. Free Press.

Anderson, P. (1999). Complexity theory and organization science. *Organization Science, 10*(3), 216–232. https://doi.org/10.1287/orsc.10.3.216

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99–120. https://doi.org/10.1177/014920639101700108

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Gandini, A. (2016). Digital work: Self-branding and social capital in the freelance knowledge economy. *Marketing Theory, 16*(1), 123–141. https://doi.org/10.1177/1470593115607942

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success. *Personnel Psychology, 58*(2), 367–408. https://doi.org/10.1111/j.1744-6570.2005.00515.x

Peters, T. (1997). The brand called you. *Fast Company, 10*, 83–90.

Seibert, S. E., Kraimer, M. L., & Liden, R. C. (2001). A social capital theory of career success. *Academy of Management Journal, 44*(2), 219–237. https://doi.org/10.2307/3069452

Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*(3), 289–310. https://doi.org/10.1214/10-STS330

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review, 20*(3), 571–610. https://doi.org/10.5465/amr.1995.9508080331

Uhl-Bien, M., Marion, R., & McKelvey, B. (2007). Complexity leadership theory. *The Leadership Quarterly, 18*(4), 298–318. https://doi.org/10.1016/j.leaqua.2007.04.002

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3–28. https://doi.org/10.1257/jep.28.2.3