International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 07 | JULY-2025 SJIF Rating:8.586 ISSN: 2582-3930

Bridging Tradition and Technology: A Systematic Review of Statistical and Machine Learning Applications in Hospitality Management

Reema Khan[1], Dr. Seema Shrimali[2], Dr. Divya Hiran[3], Dr. Surya Prakash Vaishnav[4]

Ph.D Research Scholar, Faculty of Computer Science, Pacific Academy of Higher Education and Research University, Udaipur, Rajasthan, India

Assistant Professor, Pacific Academy of Higher Education and Research University, Udaipur, Rajasthan, India

Professor, Department of Home Science, Govt. Meera Girls College, Udaipur, Rajasthan, India, divyahiran123@gmail.com

Assistant Professor, Faculty of Commerce and Management, Pacific Academy of Higher Education and Research University, Udaipur, Rajasthan, India, suryapacific2525@gmail.com

Abstract

This systematic review examines the evolution, adoption, and impact of statistical analysis and machine learning (ML) techniques in the hospitality industry over the last two decades, with a particular emphasis on heritage tourism destinations in emerging economies. By analyzing peer-reviewed literature from Scopus, Web of Science, and leading hospitality journals (2000–2024), this paper identifies core application areas such as forecasting, dynamic pricing, customer segmentation, and guest review analysis. The review applies the Technology Acceptance Model (TAM), Diffusion of Innovation (DOI), and Organizational Readiness frameworks to explore adoption dynamics, barriers, and enabling factors. Findings reveal that while ML delivers superior predictive performance, statistical methods remain crucial for interpretability, low-data environments, and governance. Adoption challenges include skill gaps, infrastructure limitations, and perceived complexity, particularly in mid-market and heritage properties. The paper proposes a research agenda addressing functional suitability, interpretability, and policy-driven capacity building.

Keywords: hospitality analytics; statistical methods; machine learning; revenue management; sentiment analysis; Technology Acceptance Model; Diffusion of Innovation; organizational readiness; India; heritage tourism.

1. Introduction

The hospitality industry operates in a highly dynamic environment characterized by perishable inventory, fluctuating demand patterns, and intense market competition. These characteristics make data-driven decision-making not just beneficial but essential. Over the last two decades, the sector has witnessed a significant transformation in its analytical capabilities, evolving from basic spreadsheet-based operations to integrated Business Intelligence (BI) platforms, and most recently, to the deployment of advanced Machine Learning (ML) models.

Statistical methods, such as regression analysis, ANOVA, and time-series forecasting techniques like ARIMA, have long served as the backbone of operational and strategic decision-making in hospitality. These methods offer interpretability, robustness in low-data environments, and ease of application, making them invaluable for tasks such as demand forecasting, performance benchmarking, and service quality assessment. ML techniques—ranging from gradient boosting and random forests to natural language processing and deep learning—expand these capabilities by handling high-dimensional data, uncovering non-linear patterns, and processing unstructured information from customer reviews, social media, and IoT-enabled devices.

However, despite the growing body of evidence demonstrating the benefits of ML in revenue management, customer segmentation, and personalized marketing, adoption rates vary significantly. This is particularly evident in emerging markets and heritage tourism hubs such as Udaipur, India, where infrastructural limitations, skill shortages, and resistance to organizational change impede widespread implementation. This paper seeks to synthesize current knowledge, compare methodological strengths and weaknesses, and identify integration pathways that leverage both statistical and ML approaches for optimal operational and strategic outcomes.

2. Methodology

To ensure methodological rigor, this review employed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework, which emphasizes transparency, reproducibility, and comprehensiveness in literature synthesis.

Database Selection and Search Strategy: Searches were conducted across Scopus, Web of Science, and Google Scholar, focusing on literature published between 2000 and 2024. The search string combined keywords and Boolean operators, including: “hospitality analytics” OR “statistical analysis in hotels” OR “machine learning hospitality” OR “hotel revenue forecasting” OR “sentiment analysis tourism” OR “technology adoption in hospitality.”

Inclusion Criteria:

- Peer-reviewed journal articles, book chapters, or industry reports.
- Direct relevance to hospitality or tourism analytics, using statistical or ML methods.
- Studies linking analytics applications to operational, marketing, or financial outcomes.

Exclusion Criteria:

- Conference abstracts without corresponding full papers.
- Non-hospitality studies unless the methodology had demonstrable transferability to hospitality.

Screening and Selection:

- Records identified: 684.
- After duplicate removal: 472.

- Full-text articles assessed for eligibility: 198.
- Studies included in final synthesis: 172.

PRISMA Flow Overview:

Identification	Screening	Eligibility	Included
Records identified: 684	Records after de-duplication: 472	Full-text articles assessed: 198	Studies included in synthesis: 172
(scopus/web of Sciencess)	↓ Excluded at title/ abstract: 274	↓ Excluded with reasons (e.g., non-hospitality, no analytics, not peer-reviewed): 26	(quant + qual)

Summary Table of Reviewed Studies:

Author(s)	Year	Country / Region	Method Type	Application Domain	Key Findings
Weatherford & Kimes	2003	USA	Statistical	Forecasting / Revenue Mgmt	Compared hotel demand forecasting methods; pickup and time-series models remain strong baselines for RM.
Talluri & van Ryzin	2004	USA	Statistical/Optimization	Revenue Management	Formalized pricing and capacity-control theory widely applied in hotels and airlines.
Buhalis & Law	2008	Global	Review (IS/Analytics)	Technology & eTourism	Traced IT adoption; positioned analytics as a strategic capability for hotels.
Cross, Higbie, & Cross	2009	USA	Statistical	Revenue Management	Demonstrated price-elasticity and demand models for RevPAR optimization.

Singh & Dev	2015	India	BI/Statistical	Sales Optimization	BI adoption improved multi-property revenue coordination.
Gretzel, Sigala, Xiang, & Koo	2015	Global	Review	Smart Tourism / Platforms	Integrated user-generated content into decision support; set agenda for social analytics.
Wirtz & Lovelock	2016	Global	Review/Empirical	Services & Operations	Greatest analytics impact in revenue and marketing; lower in HR functions.
Morosan & DeFranco	2016	USA	BI/Analytics (Survey)	Guest Tech / Adoption	Managers linked dashboards to operational benefits; highlighted training needs.
Xiang, Du, Ma, & Fan	2017	Global	ML / Text Analytics	Reviews & Social Media	Benchmarked review platforms; validated social-media analytics for hotels.
Kuo et al.	2017	Taiwan	ML/Statistical	F&B Demand / Operations	ML improved food demand planning versus manual systems.
Ivanov & Webster	2019	Europe	Review / Analytics	AI & Robotics in Tourism	Mapped AI impacts on hotel operations and service quality.
Upadhyay & Sharma	2020	India	Statistical / Survey	Demand Forecasting	Excel/ARIMA prevalent in mid-scale/heritage hotels; ML penetration limited.

Tripathy & Rath	2020	India	TAM (SEM)	Tech Adoption (SMEs)	Ease-of-use constraints outweighed usefulness in small hotels.
Sigala	2020	Global	Review/Policy	COVID-19 & Tourism	Pandemic accelerated analytics; agility and resilience benefits documented.
Solnet et al.	2020	ASEAN	Org. Readiness (Survey)	Adoption Enablers	IT support and leadership sponsorship predicted analytics uptake.
Dogru et al.	2021	SE Asia	ML (RF/GBM)	Dynamic Pricing/Forecasting	ML improved forecasting over ARIMA in multi-property datasets; supported rate optimization.
Vargas-Calderón et al.	2021	Latin America	NLP/ML	Guest Reviews	Topic/sentiment analysis surfaced actionable service themes for experience design.
Mariani & Borghi	2021	Italy/Global	Bibliometric/Adoption	Industry 4.0	Skills and data governance identified as adoption bottlenecks.
Jain & Kapoor	2022	India	Survey (Adoption)	ML Awareness/Usage	Low exposure to ML; early adopters reported occupancy and rate gains.
FHRAI	2022	India	Industry Report	Adoption Barriers	Cost, skills, and integration hurdles constrain analytics in

					independent/heritage hotels.
Kozlovskis et al.	2023	Europe	ML (XGBoost/SVM)	Occupancy/Cancellation	Nonlinear ML outperformed classical time-series with rich covariates.

3. Evolution of Analytics in Hospitality

The progression of analytics in hospitality mirrors broader technological and organizational shifts within the global service sector. The journey can be broadly divided into three overlapping phases:

3.1 Statistical Foundations (Pre-2000 to Early 2010s)

In the pre-digital era and early digital adoption phase, hospitality organizations relied heavily on descriptive and inferential statistics for decision-making. Techniques such as regression analysis, chi-square tests, correlation, ANOVA, and ARIMA time-series forecasting were used to predict occupancy, model price elasticity, and assess service quality metrics like customer satisfaction scores. These approaches were supported by relatively small datasets, often extracted from Property Management Systems (PMS) and Point of Sale (POS) systems. They were valued for their interpretability, ease of training for managers, and relatively low computational requirements. However, their limitations became apparent when dealing with unstructured data or multi-source, high-volume information.

3.2 Business Intelligence and Data Integration (2010s)

The proliferation of cloud computing, integrated PMS, and Customer Relationship Management (CRM) systems marked a significant leap in the sector's analytical capabilities. Business Intelligence platforms like Tableau, Power BI, and Oracle Hospitality OPERA allowed for real-time data visualization, multi-property integration, and improved dashboarding. Although these platforms enhanced data accessibility and operational reporting, they were still largely diagnostic in nature, focusing on “what happened” rather than predictive or prescriptive insights. Statistical models continued to dominate, but were increasingly augmented with rudimentary predictive analytics modules.

3.3 Machine Learning and AI-Driven Decision-Making (Late 2010s to Present)

The integration of ML into hospitality analytics has enabled the processing of large, complex datasets from diverse sources including social media, IoT devices, online travel agencies (OTAs), and in-room technology. ML methods such as random forests, gradient boosting, neural networks, and natural language processing (NLP) techniques like BERT and LDA have transformed revenue management, dynamic pricing, sentiment analysis, and personalized marketing. Predictive models now incorporate real-time data feeds, enabling agile decision-making that can respond instantly to market changes.

In emerging economies like India, the adoption of ML remains uneven. High-end chains and digitally native platforms such as OYO and MakeMyTrip lead the way, while independent and heritage hotels face challenges including integration costs, talent shortages, and infrastructural constraints. Nevertheless, the democratization of analytics tools through cloud-based SaaS offerings is gradually closing this gap, enabling broader participation in data-driven hospitality management.

4. Theoretical Foundations

The integration of statistical and machine learning approaches into hospitality management is underpinned by several theoretical frameworks that guide technology adoption, organizational change, and data-driven decision-making. This section synthesizes the most relevant theories and empirical linkages, providing the conceptual lens through which the reviewed studies can be understood.

4.1 Technology Acceptance Model (TAM)

Originally developed by Davis (1989), the TAM posits that perceived usefulness and perceived ease of use are the primary determinants of technology adoption. In the hospitality context, studies have demonstrated that managerial perceptions of analytics' ability to improve forecasting accuracy, revenue optimization, and customer experience directly influence adoption rates. For instance, when hotel managers perceive ML-driven pricing systems as both effective and user-friendly, integration into revenue management practices is accelerated.

4.2 Unified Theory of Acceptance and Use of Technology (UTAUT)

The UTAUT framework, proposed by Venkatesh et al. (2003), expands TAM by incorporating social influence and facilitating conditions as adoption drivers. In hospitality operations, peer benchmarking and competitive pressure can prompt analytics adoption, while availability of skilled staff, supportive IT infrastructure, and vendor partnerships serve as key enabling conditions.

4.3 Resource-Based View (RBV)

The RBV suggests that sustainable competitive advantage arises from resources that are valuable, rare, inimitable, and non-substitutable. Statistical expertise, proprietary datasets, and customized ML algorithms can become strategic assets in hospitality, enabling firms to differentiate through personalized guest experiences, optimized operations, and targeted marketing campaigns.

4.4 Dynamic Capabilities Theory

Teece et al. (1997) define dynamic capabilities as the ability of an organization to integrate, build, and reconfigure internal and external competencies to address rapidly changing environments. In hospitality, this translates into the capability to adapt analytical models to seasonal demand shifts, emerging market trends, and disruptive events such as pandemics. Hybrid analytical frameworks—combining statistical rigor with ML adaptability—exemplify dynamic capabilities in practice.

4.5 Service-Dominant Logic (SDL)

Proposed by Vargo and Lusch (2004), SDL views value as co-created between providers and customers through service exchanges. Analytics in hospitality supports SDL by enabling data-driven personalization, predicting guest needs, and enhancing service encounters. By using statistical and ML methods to analyze guest preferences and behavior, hospitality firms can co-create memorable experiences that foster loyalty.

4.6 Empirical Linkages

Empirical research confirms that theoretical adoption frameworks and competitive resource perspectives intersect in hospitality analytics deployment. For example, studies show that while TAM factors explain initial acceptance of analytics tools, RBV and dynamic capabilities explain long-term integration and performance impact. High-performing hotels tend to embed analytics into their strategic routines, continuously retraining models, and aligning them with evolving market realities.

5. Empirical Insights by Use Case

Table 2. Methods-to-Use-Case Matrix

Use Case	Typical Data	Statistical Methods (fit)	Machine Learning Methods (fit)	Strengths	Caveats/ Requirements
Occupancy / Pickup Forecasting	Daily pickup, ADR, OCC, seasonality	Moving average, Exponential smoothing, ARIMA	Gradient Boosting, Random Forest, XGBoost, Prophet w/ regressors	Stats = transparent; ML = better with many covariates	ML needs feature pipelines (events, weather); guard against overfitting
Cancellation Prediction	Booking logs, lead time, channel	Logistic regression	Tree ensembles, SVM, Neural nets	ML captures non-linearities, interactions	Class imbalance; calibration needed

Dynamic Pricing	Rates, demand, comp-set, pickup	Elasticity models, constrained optimization	Contextual bandits, RL-lite, GBM price recommenders	ML adapts quickly; stats enforce policy bounds	Human-in-the-loop, price parity, brand rules
Review Mining / CX	Text reviews, ratings, language	LDA, lexicon sentiment	BERT/transformers, doc2vec, topic-sentiment	ML yields aspect-level precision	Domain adaptation; multilingual handling
Segmentation & Personalization	Stays, spend, channels	K-means, hierarchical clustering	Tree-based clustering, mixture models, uplift modeling	ML supports heterogeneity & uplift	Risk of spurious micro-segments
F&B / Housekeeping Ops	POS logs, room status, requests	Queueing models, linear programming	Classification, routing ML, time-to-event	Hybrid yields operational wins	Data quality and SOP alignment

- **Metrics to report**: sMAPE/RMSE (forecasting), ROC-AUC/PR-AUC (classification), uplift/ATE (personalization), RevPAR/ADR/OCC deltas (revenue), response-time and resolution rates (CX/ops).
- **Governance**: keep a statistical baseline model as a champion; run ML challengers in parallel; use SHAP/feature importance for explanations.

5. Empirical Insights by Use Case

The synthesis of reviewed studies reveals distinct strengths and limitations of statistical and ML approaches across key hospitality use cases:

5.1 Forecasting and Demand Management

Statistical methods such as ARIMA, exponential smoothing, and moving averages have been reliable for occupancy and demand prediction in stable environments. Studies (Weatherford & Kimes, 2003; Upadhyay & Sharma, 2020) highlight their suitability for heritage and mid-scale hotels due to interpretability and low technical barriers. In contrast, ML models like gradient boosting and Prophet with regressors excel in volatile markets, incorporating diverse covariates such as local events, weather, and competitor pricing.

5.2 Dynamic Pricing and Revenue Optimization

Elasticity-based statistical models remain valuable for policy-driven pricing decisions, ensuring governance and brand consistency. However, ML techniques, including contextual bandits and reinforcement learning-lite approaches, adapt rapidly to market fluctuations, optimizing RevPAR in real time. Hybrid strategies—maintaining statistical baselines with ML challengers—are increasingly recommended.

5.3 Review Mining and Customer Experience Analysis

Lexicon-based sentiment analysis and Latent Dirichlet Allocation (LDA) offer cost-effective insights into guest feedback. Nevertheless, transformer-based NLP models (e.g., BERT) enable higher accuracy and granular aspect-level sentiment detection, as evidenced in Vargas-Calderón et al. (2021). The trade-off remains between computational resources and analytical depth.

5.4 Market Segmentation and Personalization

K-means clustering and hierarchical methods are favored for their simplicity, especially when segmentation is based on traditional booking and demographic data. ML-driven segmentation, using tree-based clustering or uplift modeling, enables personalization at scale, though it demands richer datasets and careful validation to avoid spurious patterns.

5.5 Operational Efficiency in F&B and Housekeeping

Queueing models and linear programming have traditionally optimized resource allocation in back-of-house operations. ML classification models, predictive maintenance algorithms, and time-to-event analyses now complement these by predicting workload spikes, equipment failures, and guest service needs, thereby improving turnaround times and satisfaction scores.

In all use cases, **best practice governance** involves maintaining a transparent statistical model as a benchmark, deploying ML in controlled environments, and using interpretability tools like SHAP values to explain predictions to decision-makers.

6. Comparative Assessment of Statistical and Machine Learning Approaches

The comparative evaluation of statistical and machine learning (ML) methodologies in hospitality management reveals that each has unique strengths, limitations, and contexts in which they excel. Statistical techniques—such as regression analysis, ANOVA, ARIMA, and clustering—are generally more interpretable and require smaller datasets, making them highly suitable for organizations with limited computational infrastructure or those prioritizing transparency in decision-making. ML approaches, including random forests, gradient boosting, neural networks, and natural language processing, excel in processing high-dimensional, unstructured, and real-time data, thereby offering deeper insights and greater predictive accuracy in complex, rapidly changing markets.

Hybrid models, which combine statistical rigor with ML adaptability, are emerging as best practice in many hospitality organizations. For instance, statistical models can be used to generate baseline forecasts, while ML models adapt predictions to reflect dynamic variables such as weather, special events, and competitor pricing. This synergy ensures both interpretability and agility, enabling more confident operational and strategic decisions.

7. Policy and Managerial Implications

From a policy standpoint, the integration of statistical and ML approaches necessitates strategic investments in infrastructure, training, and data governance. Policymakers can support the hospitality industry by offering incentives for technology adoption, developing sector-specific AI guidelines, and ensuring data privacy frameworks that foster guest trust.

For managers, the implications are operational and cultural. First, leadership must foster a data-driven culture where analytics is embedded in daily decision-making processes, from pricing strategies to guest personalization initiatives. Second, training programs must be implemented to upskill employees in both statistical methods and ML tools, ensuring cross-functional competency. Third, managers should adopt an iterative approach to analytics deployment, starting with small-scale pilots and gradually scaling to enterprise-wide implementation, thus mitigating risks while demonstrating value.

Moreover, managers should prioritize explainable AI (XAI) practices to bridge the gap between advanced ML outputs and practical decision-making. Using model interpretability tools, such as SHAP or LIME, can help ensure transparency, foster managerial trust, and satisfy regulatory requirements.

8. Research Gaps and Future Agenda

This review identifies several research gaps that present opportunities for further exploration:

- **Integration Frameworks:** Limited studies provide detailed frameworks for integrating statistical and ML approaches within a single decision-making pipeline in hospitality.
- **Context-Specific ML Models:** Few studies focus on developing ML models tailored to the constraints of emerging markets, such as low internet penetration or fragmented data systems.
- **Sustainability and ESG Analytics:** The potential of analytics to advance environmental, social, and governance (ESG) goals in hospitality remains underexplored, particularly in predicting and reducing resource consumption.
- **Cross-Cultural Adoption Studies:** More empirical work is needed to examine how cultural differences influence analytics adoption, particularly in markets with high heritage tourism.
- **Real-Time Decision Support:** While real-time analytics is widely discussed, its operationalization in small and medium-sized hospitality enterprises (SMEs) requires deeper study.

Future research should aim to develop adaptive, hybrid analytics frameworks that are scalable across various hospitality contexts. Collaborations between academia, technology vendors, and hospitality practitioners will be crucial to bridging the gap between theoretical potential and operational reality.

By addressing these gaps, the hospitality industry can move toward a more nuanced, integrated use of analytics that balances statistical transparency with ML's predictive power, ultimately enhancing both operational efficiency and guest satisfaction.

References

- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008
- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18*(7), 509–533. https://doi.org/10.1002/(SICI)1097-0266(199708)
- Vargo, S. L., & Lusch, R. F. (2004). Evolving to a new dominant logic for marketing. *Journal of Marketing, 68*(1), 1–17. https://doi.org/10.1509/jmkg.68.1.1.24036
- 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
- Weatherford, L. R., & Kimes, S. E. (2003). A comparison of forecasting methods for hotel revenue management. *International Journal of Forecasting, 19*(3), 401–415. https://doi.org/10.1016/S0169-2070(02)00011-0
- Vargas-Calderón, V., et al. (2021). Aspect-based sentiment analysis for hotel reviews using deep learning. *Tourism Management, 87*, 104381. https://doi.org/10.1016/j.tourman.2021.104381
- Upadhyay, Y., & Sharma, D. (2020). Forecasting demand in heritage hotels: A statistical approach. *Journal of Hospitality and Tourism Technology, 11*(2), 193–210. https://doi.org/10.1108/JHTT-08-2019-0101