



Data-Driven Customer Satisfaction Analysis in the Hospitality Industry: Leveraging Machine Learning to Identify Key Drivers

Hamed Fallah Tafti * 

Department of Management, Science and Arts
University, Yazd, Iran

Abstract

In the competitive and dynamic hospitality industry, customer satisfaction is pivotal to sustainable success, with online reviews on platforms like TripAdvisor providing valuable insights into guest expectations and preferences. This study aimed to uncover hidden patterns in online feedback and provide evidence-based strategies to optimize guest experiences. Using data from TripAdvisor reviews, the study employed web scraping techniques to extract textual content and numerical ratings for cleanliness, comfort, and staff service. Following data preprocessing, which involved removing missing values and handling outliers, Pearson correlation analysis and a random forest regression model were applied to assess feature significance and nonlinear relationships among variables, with cross-validation, and mean squared error used to evaluate model performance. Analysis of 2,411 reviews revealed that room quality and staff service were the most influential factors in customer satisfaction, while location and price also had significant impacts; secondary features such as views and

* Corresponding Author: h.fallah@sau.ac.ir

How to Cite: Fallah Tafti, H. (2026). Data-Driven Customer Satisfaction Analysis in the Hospitality Industry: Leveraging Machine Learning to Identify Key Drivers. (e20091). *Tourism Management Studies*, 21(74), 205 - 248. e20091 <https://doi.org/10.22054/tms.2025.88434.3114>

amenities were particularly important for specific customer segments. By integrating statistical analysis and machine learning, this research provided a comprehensive framework for understanding the drivers of customer satisfaction, emphasizing the need to prioritize room quality and staff service to enhance guest experiences. This data-driven approach offered practical strategies for optimal resource allocation and targeted marketing, fostering service innovation and strengthening competitive advantage.

1. Introduction

The hospitality industry operates within a dynamic, highly competitive global market where understanding customer satisfaction is essential for success (Wang et al., 2021). Online customer reviews, particularly on platforms like TripAdvisor, have become critical in shaping consumer decisions and hotel selection, functioning as electronic word-of-mouth (eWOM) that influences travelers' expectations (Zhang & Kim, 2021; Alsayat, 2023). Analyzing these reviews is a strategic necessity for hotels aiming to foster loyalty and profitability (Amat-Lefort et al., 2023). The advent of 'big social data' from online reviews offers opportunities for data-driven decision-making, enabling hotels to uncover insights into service quality and customer needs (Sahebi et al., 2022). Traditional methods, such as surveys, often fail to capture the nuanced, spontaneous nature of guest experiences due to their limited scope and potential bias (Wang et al., 2021). In contrast, advanced analytics, including data mining and machine learning, provide a transformative approach to processing unstructured text data and identifying key service attributes that drive satisfaction (Sharma et al., 2021). This study addresses research gaps by employing a data-driven methodology to analyze TripAdvisor reviews, focusing on identifying and quantifying the impact of service attributes on customer satisfaction in Yazd, Iran, to provide actionable recommendations for hotel management.

2. Research Methodology

This study integrates machine learning with big social data from TripAdvisor to analyze customer feedback in the hospitality sector. The methodology includes data collection, preprocessing, feature

engineering, and analysis to examine the drivers of satisfaction systematically.

Data Collection: Data were sourced from TripAdvisor due to its extensive review volume and influence (Alsayat, 2023). Using Python-based web scraping, 2,411 reviews were extracted from the top 10 hotels in Yazd, Iran, based on review volume. Extracted data included numerical ratings, textual reviews, review dates, and user identities. Reviews were translated into English, and the dataset was structured as a CSV with binary nominal variables (0 or 1) for hotel attributes and a 1-to-5 point target variable (overall rating).

Preprocessing: Missing values were removed to ensure data integrity and avoid bias from imputation (Shrestha et al., 2024). Outliers were identified and excluded using the interquartile range method. Text mining extracted hotel attributes, which were then converted into machine-learning-compatible formats. Feature engineering selected relevant attributes to optimize predictive models (Van Leeuwen & Koole, 2022).

Analysis: Descriptive statistics summarized dataset characteristics, including means and standard deviations. Pearson correlation analysis explored linear relationships between attributes and satisfaction scores. Multiple regression quantified the impacts of attributes while controlling for multiple variables (Barzizza et al., 2024). A Random Forest regression model, chosen for its robustness to high-dimensional data, was trained on a 70% training set and 30% test set, with 10-fold cross-validation to reduce overfitting (Salman et al., 2024). Model performance was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 .

3. Findings

Analysis of 2,411 TripAdvisor reviews revealed a mean overall hotel rating of 4.26 (SD = 1.00), indicating high satisfaction with variability for differentiation. Key attributes—cleanliness, comfort, location, staff service, and value—showed higher means and variability, reflecting their prominence.

Correlation Analysis: Pearson correlation identified strong relationships, with “room quality” ($r = 0.85$) and “staff service” ($r =$

0.82) as top satisfaction drivers, followed by “location” ($r = 0.75$) and “price” ($r = 0.65$).

Random Forest Model: The model ($n_estimators=100$, $random_state=42$) achieved an MSE of 1.01778, indicating reliable accuracy. Feature importance scores ranked “staff service” (0.150341), “room view” (0.124541), and “comfort” (0.094198) as the top predictors, highlighting nuanced non-linear contributions.

4. Discussion

The convergence of correlation and Random Forest results confirms that room quality and staff service are primary satisfaction drivers, aligning with the Kano model's “basic qualities,” where deficiencies cause dissatisfaction (Wu & Riantama, 2022). Comfort and amenities, as “performance qualities,” linearly enhance satisfaction and offer differentiation (Barzizza et al., 2024). Features like room view, as “attractive qualities,” delight specific segments (Nam et al., 2021). The Random Forest model's ability to capture non-linear interactions provides deeper insights, addressing complexity through explainable machine learning (Sharma et al., 2021).

5. Conclusion

This study advanced hospitality research by leveraging machine learning to analyze TripAdvisor reviews and identify room quality and staff service as key drivers of satisfaction. Recommendations include investing in room renovations, staff training, and real-time feedback monitoring using machine learning tools (Alsayat, 2023). Tailored marketing highlighting high-impact attributes and optimized pricing can strengthen competitiveness (Tan et al., 2025). These strategies align with evolving customer expectations, enhancing service quality and market positioning.

Keywords: Customer Satisfaction, Machine Learning, Online Reviews, Random Forest, Hospitality Industry



تحلیل داده‌محور رضایت مشتری در صنعت هتلداری: بهره‌گیری از یادگیری ماشین برای شناسایی محرک‌های کلیدی

حامد فلاح تفتی*  گروه مدیریت، دانشگاه علم و هنر، یزد، ایران

چکیده

در محیط رقابتی و پویای صنعت هتلداری، رضایت مشتری نقشی محوری در موفقیت پایدار ایفا می‌کند و نظرات برخط مشتریان در سکوهایی مانند تریپ‌ادوایزر به منبعی ارزشمند برای درک انتظارات و ترجیحات مهمانان تبدیل شده است. هدف این پژوهش، کشف الگوهای پنهان در بازخوردهای برخط و ارائه راهکارهای مبتنی بر شواهد برای بهینه‌سازی تجربه مهمان بود. این مطالعه با استفاده از داده‌های نظرات برخط جمع‌آوری شده از سکوی تریپ‌ادوایزر، از طریق تکنیک وب‌اسکرپینگ، داده‌های متنی و امتیازات عددی مرتبط با نظافت، راحتی و خدمات کارکنان را استخراج کرد. پس از پیش‌پردازش داده‌ها، شامل حذف مقادیر گم‌شده و مدیریت داده‌های پرت، از تحلیل همبستگی پیرسون و مدل رگرسیون جنگل تصادفی برای شناسایی اهمیت ویژگی‌ها و روابط غیرخطی میان متغیرها استفاده شد و اعتبارسنجی متقاطع با معیار میانگین مربع خطا برای ارزیابی عملکرد مدل به کار رفت. تحلیل داده‌های ۲۴۱۱ نظر نشان داد که کیفیت اتاق و خدمات کارکنان مهم‌ترین عوامل تأثیرگذار بر رضایت مشتری بودند و ویژگی‌هایی مانند موقعیت مکانی و قیمت نیز تأثیر قابل توجهی داشتند، در حالی که ویژگی‌های سطح دوم مانند منظره و امکانات رفاهی برای برخی مشتریان خاص اهمیت بیشتری داشتند. این پژوهش با ترکیب تحلیل‌های آماری و یادگیری ماشین، چارچوبی جامع برای درک محرک‌های رضایت مشتری ارائه کرد و بر لزوم تمرکز بر کیفیت اتاق و خدمات کارکنان برای بهبود تجربه مهمان تأکید داشت. این رویکرد داده‌محور، راهکارهای عملی برای تخصیص بهینه منابع و بازاریابی هدفمند پیشنهاد کرد و امکان نوآوری در خدمات و تقویت مزیت رقابتی را فراهم ساخت.

کلیدواژه‌ها: رضایت مشتری، یادگیری ماشین، نظرات برخط، جنگل تصادفی، صنعت هتلداری.

مقدمه

صنعت مهمان‌نوازی امروز در بازارهای جهانی، پویا و به شدت رقابتی فعالیت می‌کند؛ بازاری که در آن درک و رضایت مشتری برای موفقیت پایدار حیاتی است (Wang et al., 2021). در چنین محیطی، نظرات برخط مشتریان فراتر از یک سازوکار بازخورد ساده عمل کرده و به منبعی کلیدی برای تصمیم‌گیری مصرف‌کننده و انتخاب هتل بدل شده‌اند (Zhang & Kim, 2021). مسافران بیش‌ازپیش به تجربه‌های هم‌تایان خود در سکوهایی مانند تریپ‌ادوایزر تکیه می‌کنند و این روایت‌های دیجیتال را شاخصی معتبر برای کیفیت خدمات و ارزش تلقی می‌کنند (Wang et al., 2021). این پدیده که به «دهان‌به‌دهان الکترونیکی»^۱ شناخته می‌شود، به‌طور معناداری بر انتخاب‌های آینده و شکل‌گیری انتظارات مشتری اثر می‌گذارد (Alsayat, 2023)؛ بنابراین برای هتل‌هایی که به دنبال پرورش وفاداری و افزایش سودآوری هستند، شنیدن فعال و تفسیر این مخازن عظیم بازخورد، ضرورتی راهبردی است (Amat-Lefort et al., 2023).

حجم و شتاب بالای این نظرات که با عنوان «داده‌های بزرگ اجتماعی»^۲ شناخته می‌شوند، چالش و فرصتی منحصربه‌فرد برای تصمیم‌سازی داده‌محور در مهمان‌نوازی فراهم می‌آورد (Sahebi et al., 2022 ; Alsayat, 2023). در بازار به شدت رقابتی امروز، شناخت دقیق ترجیحات و انتظارات مسافران دیگر یک مزیت لوکس نیست؛ بلکه پیش‌نیازی برای برنامه‌ریزی راهبردی و برتری عملیاتی است (Diwan, 2025). با واکاوی این ردپاهای دیجیتال، مدیران می‌توانند دیدگاه‌های حیاتی درباره کیفیت خدمات کسب کنند، حوزه‌های بهبود را بیابند و نیازهای برآورده نشده را کشف نمایند و بدین سان جایگاه رقابتی خود را تقویت کنند (Nam et al., 2021). این رویکرد، فراتر از شواهد روایتی رفته و مبنایی تجربی برای بازطراحی پیشنهادهای خدمتی و بهینه‌سازی تجربه مهمان ارائه می‌دهد (Sahebi et al., 2022).

1. Electronic word-of-mouth - eWOM

2. Social Big Data

به صورت تاریخی هتل‌ها برای سنجش رضایت مشتری به کارت‌های نظرسنجی، پرسشنامه‌ها یا گروه‌های کانونی متکی بوده‌اند. هرچند این ابزارها ارزشمندند، اما از محدودیت‌هایی مانند دامنه محدود سنجش، سوگیری بالقوه و ناتوانی در ثبت ابعاد خودانگیخته، دقیق و بدون ساختار تجربه واقعی مهمان رنج می‌برند (Amat-Lefort et al., 2023; Wang et al., 2021). در نتیجه، تصویری محدود ارائه می‌کنند و نه تنوع تجربه یک پایگاه بزرگ مشتری را بازمی‌تابانند و نه «نیازهای پنهان»^۱ را آشکار می‌سازند (Nam et al., 2021; Zhang & Kim, 2021). در مقابل، ظهور داده‌کاوی و الگوریتم‌های یادگیری ماشین جایگزینی دگرگون‌ساز است؛ رویکردهایی که در پردازش متن‌های انبوه نظرات برخط، استخراج دیدگاه‌های دقیق درباره ویژگی‌های خدمت و کشف روابط پیچیده و غیرخطی میان مؤلفه‌های هتل و رضایت کلی، توانا تر از روش‌های سنتی‌اند (Sharma et al., 2021; Sahebi et al., 2022).

با وجود پیشرفت‌های نظری در مدل‌های تبیینی رفتار مصرف‌کننده، مانند پارادایم انتظار/ناهمخوانی و مدل کانو که بر جنبه‌های روان‌شناختی و نامتقارن رضایت تمرکز دارند، یک شکاف نظری برجسته میان این مدل‌ها و کاربردهای عملی داده‌محور در صنعت هتلداری وجود دارد. این مدل‌ها اغلب به صورت مفهومی باقی می‌مانند و فاقد ابزارهای کمی برای پردازش داده‌های بزرگ و بدون ساختار مانند نظرات برخط هستند که این امر منجر به عدم توانایی در شناسایی روابط پیچیده و پنهان میان ویژگی‌های خدمات و رضایت واقعی مشتری می‌شود. نوآوری این پژوهش در پر کردن این شکاف نهفته است؛ با ادغام مدل‌های نظری رفتار مصرف‌کننده با تکنیک‌های پیشرفته یادگیری ماشین، مانند رگرسیون جنگل تصادفی، چارچوبی داده‌محور و یکپارچه ارائه می‌شود که نه تنها روابط خطی و غیرخطی را کمی‌سازی می‌کند، بلکه ارزش افزوده‌ای عملی برای مدیران هتل فراهم می‌آورد تا استراتژی‌های بهبود خدمات را بر اساس شواهد تجربی اولویت‌بندی کنند و بدین ترتیب، به تقویت مزیت رقابتی پایدار کمک نماید.

هزینه‌های جذب مشتری جدید به مراتب بیشتر از هزینه‌های حفظ مشتریان فعلی است؛ از این رو، شناسایی و نگهداری مشتریان وفادار از اولویت بالایی برخوردار است. تقسیم‌بندی مشتریان با استفاده از روش‌های مبتنی بر داده کاوی به سازمان‌ها کمک می‌کند تا استراتژی‌های مؤثری برای ارائه خدمات باکیفیت تدوین کنند (Ansari & Asadi, 2016). با وجود ارزش اثبات‌شده نظرات برخط و ظرفیت‌های یادگیری ماشین، هنوز شکاف معناداری در ترجمه نظام‌مند این ثروت بدون ساختار به استراتژی‌های عملی و شواهدمحور برای بهینه‌سازی خدمات هتل وجود دارد. پژوهش‌های پیشین ضمن تأیید اهمیت نظرات برخط، غالباً با محدودیت‌های روش‌شناختی مواجه بوده‌اند، بر گونه‌های خاصی از هتل‌ها تمرکز کرده یا به چارچوب‌های تحلیلی کمتر فراگیر تکیه کرده‌اند (Zhang & Kim, 2021). این مطالعه با هدف پر کردن این شکاف، از رویکردی استوار و داده‌محور بهره می‌گیرد و با تکیه بر تکنیک‌های پیشرفته یادگیری ماشین، تحلیل جامعی از اهمیت ویژگی‌ها در نظرات هتل‌های تریپ‌ادوایزر ارائه می‌کند. بدین وسیله، علاوه بر شناسایی ویژگی‌های کلیدی اثرگذار بر رضایت و امتیاز کلی، توصیه‌های عملی و شواهدمحور برای مدیریت هتل، راهبردهای بازاریابی و ابتکارهای بهبود خدمات عرضه می‌شود و بدین سان به ادبیات دانشگاهی و صنعت، دیدگاه‌های تازه‌ای می‌افزاید.

این پژوهش با هدف ارائه درک دقیق و مبتنی بر شواهد از عوامل مؤثر بر رضایت مشتری در صنعت هتلداری، به‌طور سیستماتیک به پیچیدگی‌های بازخورد برخط می‌پردازد. در حالی که ادبیات موجود بر اهمیت نظرات مشتریان تأکید دارد، شکافی در شناسایی کمی و دقیق ویژگی‌های خدماتی که بیشترین تأثیر را بر تجربه کلی مهمان دارند، وجود دارد (Nam et al., 2021). این مطالعه با تمرکز بر این خلأ، سه هدف اصلی را دنبال می‌کند که به‌طور مستقیم به بهبود استراتژی‌های عملیاتی و بازاریابی هتل‌ها کمک می‌کند.

هدف نخست، شناسایی مهم‌ترین ویژگی‌های خدمات هتل است که به‌طور قابل توجهی بر رضایت مشتری و امتیازات کلی تأثیر می‌گذارند. این امر مستلزم فراتر رفتن از همبستگی‌های ساده و درک «تأثیر واقعی» بهبود خدمات بر رضایت کلی است.

(Mirthipati, 2024). هدف این است که نه تنها عوامل مؤثر را فهرست کنیم، بلکه «تأثیر کمی» هر یک از این ویژگی‌ها را بر رضایت مشتری مشخص نماییم و بدین ترتیب به مدیران هتل امکان دهیم تا تلاش‌های خود را بر مهم‌ترین حوزه‌ها متمرکز کنند (Sharma et al., 2021; Barzizza et al., 2024).

هدف دوم، توسعه و به‌کارگیری یک روش شناسی قوی و مبتنی بر داده‌ها برای تحلیل نظرات برخط مشتریان است. این روش شناسی با ادغام تکنیک‌های پیشرفته تحلیلی، قادر به «کشف الگوهای پنهان و ارائه دیدگاه‌های ارزشمند» از داده‌های متنی بدون ساختار است (Zhang & Kim, 2021). این رویکرد به دنبال ارائه «تحلیل دقیق و پالایش‌شده‌ای از عوامل کیفیت خدمات» است که از طریق چارچوب‌های تحلیل رضایت مشتری با استفاده از یادگیری ماشین حاصل می‌شود (Diwan, 2025; Barzizza et al., 2024). این امر شامل شناسایی «اجزای ارزشی» از دیدگاه مشتری با استفاده از هوش مصنوعی است (Sahebi et al., 2022).

سومین هدف، ارائه دیدگاه‌هایی عملی و مبتنی بر شواهد برای مدیریت هتل، استراتژی‌های بازاریابی و ابتکارات بهبود خدمات است. این نگاه‌ها به هتل‌ها کمک می‌کند تا «تأثیر موفقیت‌آمیز بهبود ویژگی‌های خاص محصول بر رضایت مشتری» را پیش‌بینی کنند (Barzizza et al., 2024). هدف نهایی، ارائه «توصیه‌های عملی» است که به «توسعه استراتژیک پایدار و مدیریت صنعت» کمک کرده و به بهینه‌سازی تجربه مهمان و حتی «بهینه‌سازی قیمت‌گذاری محصول» منجر شود (Zhang & Kim, 2021; Tan et al., 2025).

بر اساس این اهداف، سؤالات پژوهشی زیر مطرح می‌شوند:

- کدام یک از ویژگی‌های خدمات هتل (مانند نظافت، کارکنان، امکانات) که در نظرات برخط تریپ‌ادوایزر منعکس شده‌اند، بیشترین تأثیر کمی را بر امتیازات کلی رضایت مشتری اعمال می‌کنند؟

• چگونه این ویژگی‌های خدماتی تأثیرگذار با یکدیگر تعامل دارند و چه «قوانین همبستگی» یا الگوهای ساختاری زیربنایی از بازخورد مشتری برای توضیح تغییرات در رضایت ظاهر می‌شوند؟

• چه توصیه‌های عملی و مبتنی بر شواهد را می‌توان برای مدیریت هتل و تیم‌های بازاریابی به منظور ارتقای استراتژیک کیفیت خدمات و بهینه‌سازی تجربیات مهمانان بر اساس اهمیت ویژگی‌های شناسایی شده و الگوهای تعاملی فرموله کرد؟

از منظر روش شناختی، این مطالعه با ادغام یادگیری ماشین پیشرفته با داده‌های بزرگ رسانه‌های اجتماعی، محدودیت رویکردهای سنتی پرهزینه و زمان‌بر را برطرف می‌کند (Amat-Lefort et al., 2023). کاربرد تکنیک‌هایی مانند رگرسیون جنگل تصادفی برای استخراج اهمیت ویژگی‌ها، مکانیزمی قوی، عینی و مقیاس‌پذیر برای ثبت روابط و تعاملات پیچیده و غیرخطی در بازخورد متنی ارائه می‌دهد (Zhang & Kim, 2021). این امر سرعت، دقت و گستردگی پژوهش بازاریابی را افزایش داده و هزینه‌های تحلیلی را کاهش می‌دهد (Sahebi et al., 2022).

دستاوردهای این پژوهش، مدیران هتل را قادر می‌سازد تا بهبودهای استراتژیک را اولویت‌بندی کرده و تخصیص منابع را بهینه سازند (Mirthipati, 2024; Alsayat, 2023). این دیدگاه‌های داده‌محور تیم‌های بازاریابی را برای پالایش استراتژی‌های تبلیغاتی و تصمیمات سرمایه‌گذاری با روشن کردن اینکه کدام امکانات بهترین بازده را دارند، توانمند می‌سازد (Tan et al., 2025; Amat-Lefort et al., 2023; Sharma et al., 2021).

پیشینه پژوهش

نظریه انتظار-ناهمخوانی و منطق غالب خدمت

در این پژوهش، دو چارچوب نظری مکمل، یعنی نظریه انتظار-ناهمخوانی^۱ و منطق غالب خدمت^۲، به عنوان مبنای تحلیلی اتخاذ شده‌اند. نظریه انتظار-ناهمخوانی که به طور گسترده

1. Expectation-Disconfirmation Theory (EDT)

2. Service-Dominant Logic (S-D Logic)

در مطالعات رضایت مشتری به کار رفته است، بیان می‌دارد که رضایت، نتیجه یک فرآیند مقایسه شناختی بین انتظارات پیش از مصرف مشتری و عملکرد درک شده محصول یا خدمت است (Wüst and Bremser, 2025; Wang et al., 2021). بر اساس این پارادایم، زمانی که عملکرد واقعی فراتر از انتظارات باشد، تأیید مثبت (رضایت) رخ می‌دهد و در مقابل، عملکرد ضعیف‌تر از انتظارات منجر به تأیید منفی (نارضایتی) می‌شود (Cutler and Carmichael, 2010). این نظریه به‌ویژه برای تحلیل داده‌های کمی مانند امتیازات مشتریان به ویژگی‌های خاص هتل (مانند نظافت، امکانات و خدمات کارکنان) که در آن مشتریان به طور مستقیم عملکرد را با یک استاندارد مورد انتظار مقایسه می‌کنند، بسیار مناسب است (Wu and Riantama, 2022). با این حال، پژوهش‌ها نشان داده‌اند که تمرکز صرف بر این مدل شناختی ممکن است ابعاد عاطفی و تجربی پیچیده‌تر تعاملات خدماتی را نادیده بگیرد (Prayag et al., 2013).

برای پوشش این شکاف، منطق غالب خدمت یک دیدگاه معاصر و جامع‌تر ارائه می‌دهد که از تمرکز سنتی بر «کالا» به سمت درک خدمت به‌عنوان فرآیند اصلی تبادل اقتصادی حرکت می‌کند (Fouad et al., 2024). بر خلاف دیدگاه سنتی که مشتری را یک دریافت‌کننده منفعل ارزش می‌داند، منطق غالب خدمت بر هم‌آفرینی ارزش تأکید دارد، جایی که مشتریان به‌عنوان یک عامل فعال در شکل‌دهی به تجربه خود مشارکت می‌کنند (Iloranta and Komppula, 2022). در زمینه هتلداری، این بدان معناست که ارزش تنها توسط هتل «ارائه» نمی‌شود، بلکه در طول تعاملات بین مهمان، کارکنان، سایر مهمانان و محیط فیزیکی هتل به‌طور مشترک «خلق» می‌شود (Adam, 2021). نظرات برخاسته از مشتریان، تجلی بارز این فرآیند هم‌آفرینی است؛ مشتریان از طریق به اشتراک‌گذاری تجربیات خود، نه تنها ارزش شخصی خود را محقق می‌سازند، بلکه در ساختن شهرت و تصویر برند هتل نیز مشارکت فعال دارند. در حالی که نظریه انتظار-ناهمخوانی مکانیزم ارزیابی ویژگی‌های منفرد خدمت را تبیین می‌کند، منطق غالب خدمت زمینه وسیع‌تری برای درک چگونگی شکل‌گیری تجربه کلی از طریق تعاملات پویا و هم‌آفرینی ارزش فراهم می‌سازد.

رضایت مشتری در چارچوب مدل کانو

مدل کانو یک چارچوب تحلیلی قدرتمند برای درک رابطه نامتقارن و غیرخطی بین ویژگی‌های خدمات و رضایت مشتری ارائه می‌دهد. این مدل فرض می‌کند که همه ویژگی‌های محصول یا خدمت تأثیر یکسانی بر رضایت ندارند و آن‌ها را بر اساس نحوه تأثیرگذاری بر احساسات مشتری به سه دسته اصلی طبقه‌بندی می‌کند (Sun et al., 2025). دسته اول، الزامات اساسی یا بایده‌ها هستند که مشتریان آن‌ها را بدیهی می‌دانند؛ حضور آن‌ها صرفاً از ناراضیاتی جلوگیری می‌کند، اما غیابشان به شدت منجر به ناراضیاتی می‌شود (Wang et al., 2021). به عنوان مثال، نظافت اتاق یا امنیت در یک هتل از این دسته هستند. دسته دوم، الزامات عملکردی یا تک‌بعدی هستند که رابطه‌ای خطی با رضایت دارند؛ هرچه عملکرد این ویژگی‌ها بهتر باشد، رضایت مشتری بیشتر می‌شود و بالعکس. سرعت اینترنت بی‌سیم یا کیفیت صبحانه نمونه‌هایی از این ویژگی‌ها هستند. دسته سوم، الزامات جذاب یا هیجان‌انگیز هستند که مشتریان انتظار آن‌ها را ندارند و غیابشان موجب ناراضیاتی نمی‌شود، اما حضورشان به طور نامتناسبی سطح رضایت را افزایش داده و به وجد می‌آورد (Veréb and Azevedo, 2019). ارائه یک هدیه خوشامدگویی غیرمنتظره یا ارتقای رایگان اتاق در این طبقه قرار می‌گیرند.

مطالعات تجربی متعددی در صنعت هتلداری از مدل کانو برای شناسایی نیازهای پنهان و آشکار مشتریان و اولویت‌بندی بهبود خدمات استفاده کرده‌اند. این پژوهش‌ها اغلب از طریق پرسشنامه‌های ساختاریافته، ویژگی‌های مختلف خدمات هتل را در دسته‌های کانو طبقه‌بندی کرده و تأثیر نامتقارن آن‌ها را بر رضایت کلی تأیید نموده‌اند (Wang et al., 2021). برای مثال، مشخص شده است که سرمایه‌گذاری بیش از حد بر روی ویژگی‌های اساسی (مانند بهبود نظافتی که از قبل در سطح قابل قبولی قرار دارد) بازدهی نزولی در افزایش رضایت دارد، در حالی که تمرکز بر ویژگی‌های عملکردی و معرفی هوشمندانه ویژگی‌های جذاب می‌تواند مزیت رقابتی پایدار ایجاد کند (Nam et al., 2021; Sahebi et al., 2022).

یکی از ملاحظات کلیدی در استفاده از مدل کانو، ماهیت پویای انتظارات مشتری است؛ یک ویژگی که امروز جذاب تلقی می‌شود، ممکن است در آینده به یک الزام اساسی تبدیل شود (Wu and Riantama, 2022). این پویایی، اهمیت تحلیل مستمر بازخوردهای مشتریان را دوچندان می‌کند. اگرچه پژوهش حاضر مستقیماً از روش پیمایشی کانو استفاده نمی‌کند، اما چارچوب مفهومی آن برای تفسیر نتایج حاصل از تحلیل یادگیری ماشین بسیار ارزشمند است. ویژگی‌هایی که توسط الگوریتم‌ها به عنوان عوامل تأثیرگذار قوی شناسایی می‌شوند، می‌توانند با منطق کانو تحلیل شوند تا مشخص گردد آیا آن‌ها به عنوان عوامل پیشگیرانه از نارضایتی (بایدی)، محرک‌های اصلی رضایت (عملکردی) یا عوامل ایجادکننده وفاداری و تمایز (جذاب) عمل می‌کنند. این چارچوب به پر کردن شکاف بین شناسایی آماری اهمیت ویژگی‌ها و درک روان‌شناختی تأثیر آن‌ها بر تجربه مهمان کمک شایانی می‌کند (Barzizza et al., 2024; Alsayat, 2023).

تحلیل احساسات در نظرات برخط

تحلیل احساسات^۱، به عنوان مطالعه محاسباتی نظرات، نگرش‌ها و عواطف افراد، به ابزاری حیاتی برای استخراج دیدگاه‌های مشتریان از حجم عظیم محتوای تولیدشده توسط کاربر تبدیل شده است (Medhat et al., 2014). در صنعت مهمان‌نوازی که در آن تجربیات ذهنی و عاطفی نقشی اساسی ایفا می‌کنند، نظرات برخط منبعی غنی و بدون واسطه از بازخوردهای واقعی هستند که برخلاف پیمایش‌های سنتی، تحت تأثیر اثرات آزمایشگاهی قرار نمی‌گیرند (Amat-Lefort et al., 2023). پردازش زبان طبیعی^۲ به عنوان یک حوزه از هوش مصنوعی، مجموعه‌ای از تکنیک‌های اساسی را برای تبدیل این داده‌های متنی بدون ساختار به اطلاعات ساختاریافته و قابل تحلیل فراهم می‌کند. از طریق این پردازش، می‌توان قطبیت احساسی (مثبت، منفی یا خنثی) را در سطح سند، جمله یا حتی جنبه‌های خاصی از خدمات شناسایی کرد و بدین ترتیب درک عمیق‌تری از ادراکات مهمانان به دست آورد.

1. Sentiment Analysis
2. Natural Language Processing (NLP)

رویکردهای روش شناختی در تحلیل احساسات عمدتاً به دو دسته مبتنی بر واژگان^۱ و مبتنی بر یادگیری ماشین^۲ تقسیم می‌شوند. رویکردهای مبتنی بر واژگان از فرهنگ‌های لغت از پیش تعریف‌شده‌ای استفاده می‌کنند که در آن به کلمات، امتیازات احساسی تخصیص داده شده است (Tan et al., 2025). این روش‌ها به دلیل سادگی و عدم نیاز به داده‌های آموزشی، کاربرد گسترده‌ای دارند، اما در تشخیص مفاهیم پیچیده مانند کنایه، انکار و وابستگی به زمینه، با محدودیت مواجه هستند. در مقابل، رویکردهای مبتنی بر یادگیری ماشین، با آموزش مدل‌هایی مانند ماشین‌های بردار پشتیبان^۳، بیز ساده^۴ یا شبکه‌های عصبی عمیق بر روی مجموعه داده‌های برچسب‌گذاری‌شده، به دقت بالاتری، به‌ویژه در درک زمینه‌های متنی، دست می‌یابند (Sharma et al., 2021; Alsayat, 2023). مطالعات اخیر به‌طور فزاینده‌ای به سمت مدل‌های یادگیری عمیق مانند BERT و LSTM گرایش پیدا کرده‌اند که توانایی بالایی در درک روابط معنایی پیچیده در زبان طبیعی دارند (Benaddi et al., 2024).

در ادبیات گردشگری و هتلداری، پژوهش‌های متعددی از این تکنیک‌ها برای ارزیابی تجربه مشتری بهره برده‌اند. برخی مطالعات با استفاده از تحلیل احساسات در سطح کل نظر، یک امتیاز کلی برای هر بازخورد محاسبه کرده و رابطه آن را با امتیاز ستاره‌ای بررسی نموده‌اند (Wang et al., 2021). با این حال، تحقیقات پیشرفته‌تر به سمت تحلیل احساسات مبتنی بر جنبه^۵ حرکت کرده‌اند که امکان شناسایی و ارزیابی احساسات مشتریان نسبت به ویژگی‌های خاص خدمات (مانند «کارکنان»، «اتاق» یا «صبحانه») را فراهم می‌آورد (Mehra, 2023). این رویکرد دانشی دقیق و عملیاتی در اختیار مدیران قرار می‌دهد تا نقاط قوت و ضعف خود را در سطح جزئی شناسایی کنند (Garner et al., 2022). برای مثال، یک هتل ممکن است امتیاز کلی بالایی داشته باشد، اما تحلیل مبتنی بر جنبه، نشان می‌دهد که احساسات منفی قابل توجهی در مورد فرآیند پذیرش وجود دارد.

-
1. Lexicon-based
 2. Machine Learning-based
 3. Support Vector Machine (SVM)
 4. Naive Bayes
 5. Aspect-Based Sentiment Analysis (ABSA)

یادگیری ماشین در خدمات و تحلیل اهمیت ویژگی‌ها

با افزایش حجم داده‌های تولیدشده توسط مشتریان، صنعت مهمان‌نوازی به‌طور فزاینده‌ای از روش‌های آماری سنتی به سمت الگوریتم‌های یادگیری ماشین برای تحلیل و پیش‌بینی رفتار مشتری حرکت کرده است. کاربرد این الگوریتم‌ها در بخش خدمات، دو هدف اصلی را دنبال می‌کند: پیش‌بینی و تفسیرپذیری. در حالی که مدل‌های پیش‌بینی‌کننده، مانند شبکه‌های عصبی عمیق^۱ و ماشین‌های بردار پشتیبان، در پیش‌بینی تقاضا، نرخ اشغال هتل و احتمال ریزش مشتری به‌دقت بالایی دست یافته‌اند (Diwan, 2025)، ارزش اصلی آن‌ها در توانایی مدل‌سازی روابط پیچیده و غیرخطی نهفته است که اغلب از دید تحلیل‌های رگرسیون خطی پنهان می‌ماند (Sahebi et al., 2022). با این حال، بسیاری از این مدل‌های پیچیده به‌عنوان «جعبه سیاه» عمل می‌کنند که درک چگونگی رسیدن به یک تصمیم خاص را برای مدیران دشوار می‌سازد (Sharma et al., 2021). این محدودیت، اهمیت رویکردهای متمرکز بر تفسیرپذیری را برجسته می‌سازد که هدف آن‌ها نه تنها پیش‌بینی دقیق، بلکه شناسایی و رتبه‌بندی عوامل کلیدی مؤثر بر یک نتیجه مشخص، مانند رضایت مشتری است.

در این راستا، تکنیک‌های تحلیل اهمیت ویژگی به ابزاری قدرتمند برای مدیران هتل تبدیل شده‌اند تا بتوانند منابع خود را بر روی بهبود جنبه‌هایی از خدمات متمرکز کنند که بیشترین تأثیر را بر تجربه مهمان دارند. الگوریتم‌های مبتنی بر درخت تصمیم، به‌ویژه جنگل تصادفی^۲، به دلیل توانایی ذاتی خود در ارزیابی اهمیت ویژگی‌ها، محبوبیت ویژه‌ای کسب کرده‌اند (Salman et al., 2024). جنگل تصادفی که یک مدل یادگیری گروهی^۳ است، با ساخت تعداد زیادی درخت تصمیم و تجمیع نتایج آن‌ها، نه تنها به پیش‌بینی‌های قوی دست می‌یابد، بلکه می‌تواند اهمیت هر متغیر را با اندازه‌گیری میانگین کاهش ناخالصی (مانند شاخص جینی) در تمام درختان مجموعه، محاسبه کند (Ren et al., 2024). این روش، برخلاف تحلیل رگرسیون که بر مفروضات خطی بودن و عدم هم‌خطی استوار

1. Deep Neural Network (DNN)

2. Random Forest

3. Ensemble Learning

است، قادر به شناسایی روابط غیرخطی و اثرات متقابل بین متغیرهاست؛ برای مثال، ممکن است تأثیر کیفیت صبحانه بر رضایت کلی، به موقعیت مکانی هتل وابسته باشد، رابطه‌ای که مدل‌های خطی به سادگی قادر به ثبت آن نیستند.

پژوهش‌های مختلف در حوزه صنعت گردشگری با استفاده از ابزارهای یادگیری ماشین انجام شده است تا تحلیل‌های کمی و دقیقی از وضعیت این صنعت را نشان دهند. برای مثال در مطالعاتی پیش‌بینی دقیق تعداد گردشگران ورودی به مقاصد گردشگری که مسئله‌ای حیاتی است با استفاده از داده‌ها و رفتار جست‌وجوگری برخط کاربران تحلیل شده است (Fallah Tafti, 2020).

پژوهش‌های اخیر در حوزه هتلداری به‌طور گسترده از این قابلیت بهره برده‌اند. به‌عنوان نمونه، مطالعاتی با به‌کارگیری جنگل تصادفی و سایر مدل‌های یادگیری ماشین، توانسته‌اند ویژگی‌های خدماتی مانند کیفیت تعامل کارکنان، تمیزی و نمای اتاق را به‌عنوان مهم‌ترین پیش‌بینی‌کننده‌های امتیاز نهایی مشتری شناسایی کنند (Alsayat, 2023). این رویکرد داده‌محور به کسب‌وکارها اجازه می‌دهد تا از تصمیم‌گیری‌های مبتنی بر شهود فاصله گرفته و به سمت استراتژی‌های بهبود خدمات مبتنی بر شواهد حرکت کنند (Elhabyb et al., 2024).

شکاف‌های پژوهشی و فرصت‌ها

مرور نظام‌مند ادبیات حاکی از وجود شکاف‌های مفهومی و روش‌شناختی معناداری است. یک خلأ برجسته، جدایی مشهود بین مطالعاتی است که بر پایه‌های نظری استوارند و آن‌هایی که رویکردهای داده‌محور را به‌کار می‌گیرند. پژوهش‌های مبتنی بر چارچوب‌های نظری مانند مدل کانو، اغلب به داده‌های پیمایشی متکی هستند که با محدودیت‌هایی نظیر سوگیری در پاسخ‌دهی و اندازه نمونه محدود مواجه‌اند (Sharma et al., 2021; Tan et al., 2025). از سویی دیگر، مطالعاتی که از حجم عظیم داده‌های برخط بهره می‌برند، هرچند در شناسایی الگوهای پنهان موفق عمل می‌کنند، اما غالباً فاقد یک چارچوب نظری برای تفسیر عمیق‌تر یافته‌ها هستند و صرفاً به شناسایی همبستگی‌ها اکتفا می‌کنند (Zhang

(Kim and, 2021). این دوگانگی منجر به وضعیتی شده است که در آن، غنای مفهومی نظریه‌ها به ندرت با قدرت تحلیلی روش‌های پیشرفته داده‌کاوی تلفیق می‌شود. علاوه بر این، در حوزه مطالعات مبتنی بر یادگیری ماشین، یک تنش بنیادین بین دقت پیش‌بینی و تفسیرپذیری مدیریتی وجود دارد. بسیاری از الگوریتم‌های پیچیده که دقت بالایی در پیش‌بینی رضایت مشتری ارائه می‌دهند، به مثابه یک «جعبه سیاه» عمل کرده و درک روشنی از منطق تصمیم‌گیری خود ارائه نمی‌دهند (Sharma et al., 2021). این امر کاربرد عملیاتی یافته‌ها را برای مدیران هتل که نیازمند بینش‌های شفاف و قابل اجرا برای تخصیص منابع و بهبود خدمات هستند، به شدت محدود می‌سازد. بنابراین، نیاز مبرمی به پژوهش‌هایی وجود دارد که نه تنها بر بهینه‌سازی دقت مدل تمرکز کنند، بلکه به همان اندازه بر «گشودن جعبه سیاه» و ارائه نتایج قابل تفسیر برای تصمیم‌گیرندگان تأکید ورزند (Budhwar et al., 2023).

روش

با توجه به ماهیت داده‌محور مطالعه و اهمیت بازخوردهای مشتریان در صنعت مهمان‌نوازی، نظرات برخط آن‌ها از سکوی تریپ‌ادوایزر به عنوان منبع اصلی داده انتخاب شد. تریپ‌ادوایزر به دلیل حجم وسیع نظرات، جامعیت اطلاعات و تأثیرگذاری بر تصمیم‌گیری مصرف‌کنندگان، یک بستر ایده‌آل برای استخراج دیدگاه‌های مربوط به رضایت مشتری است (Zhang & Kim, 2021 ; Nam et al. 2021). این سکو، داده‌های متنی غنی و همچنین، امتیازات عددی برای جنبه‌های مختلف خدمات هتل را فراهم می‌آورد که امکان تحلیل عمیق‌تری را نسبت به نظرسنجی‌های سنتی فراهم می‌کند (Alsayat, 2023; Amat-Lefort et al., 2023).

داده‌های مورد استفاده در این مطالعه از وب‌سایت عمومی تریپ‌ادوایزر^۱ استخراج شده‌اند، که نظرات کاربران به صورت آزاد و عمومی در دسترس قرار دارد. این داده‌ها شامل اطلاعات ناشناس (مانند متن نظرات و امتیازات عددی بدون شناسایی مستقیم

1. Tripadvisor.com

کاربران) هستند و هیچ داده شخصی حساس (مانند نام واقعی یا اطلاعات تماس) جمع‌آوری نشده است. استفاده از این داده‌ها با رعایت اصول اخلاقی پژوهش، انجام شده است؛ به‌ویژه، تمرکز بر اهداف تحقیقاتی غیرتجاری، عدم نقض حریم خصوصی و اطمینان از اینکه استخراج داده‌ها هیچ آسیبی به کاربران یا پلتفرم وارد نمی‌کند.

برای گردآوری این داده‌ها، از تکنیک خراش وب^۱ استفاده شد که به این منظور از زبان برنامه‌نویسی پایتون استفاده گردید تا به‌صورت نظام‌مند به صفحات فهرست هتل‌ها در تریپ‌ادوایزر دسترسی پیدا کند و داده‌های نظرات را استخراج نماید. فرآیند شامل جمع‌آوری اطلاعاتی نظیر امتیازات عددی برای ویژگی‌های خاص هتل، نظرات متنی مشتریان، تاریخ ثبت نظر و هویت کاربر بود (Zhang & Kim, Nam et al., 2021; 2021). همچنین، قابلیت مدیریت صفحه‌بندی^۲ در اسکرپت تعبیه شد تا اطمینان حاصل شود که تعداد زیادی از نظرات برای هر هتل جمع‌آوری می‌شود. داده‌های استخراج‌شده سپس در قالب ساختاریافته^۳ ذخیره شدند تا برای تحلیل‌های بعدی آماده گردند. با اجرای این فرآیند تعداد ۲۴۱۱ نظر از مشتریان ۱۰ هتل برتر استان یزد (بر اساس تعداد نظرات موجود در سامانه) استخراج شدند. لازم به ذکر است به‌دلیل حفظ محرمانگی و ماهیت کیفی نظرات گردآوری‌شده از ذکر نام این ۱۰ هتل خودداری شده است. انتخاب ده هتل برتر استان یزد بر اساس تعداد نظرات موجود در سامانه تریپ‌ادوایزر انجام شد که این معیار به‌عنوان شاخصی از محبوبیت و فعالیت کاربران عمل می‌کند. این رویکرد نمونه‌گیری هدفمند بر پایه حجم داده است، زیرا هتل‌هایی با بیشترین تعداد نظرات (حداقل ۲۰۰ نظر به ازای هر هتل) انتخاب شدند تا حجم کلی نمونه (۲۴۱۱ نظر) برای تحلیل‌های آماری و یادگیری ماشین کافی باشد. محاسبه اندازه نمونه بر اساس فرمول کوکران برای جوامع نامحدود (با سطح اطمینان ۹۵ درصد و حاشیه خطای ۵ درصد) نشان می‌دهد که این حجم داده، تعمیم‌پذیری یافته‌ها را تضمین می‌کند و از سوگیری ناشی از نمونه‌های

1. Web scraping

2. Pagination

3. Comma Separated Values (CSV)

کوچک جلوگیری می‌نماید. همچنین، این انتخاب تمرکز بر هتل‌های پربازدید را فراهم می‌آورد که نماینده تجربیات متنوع مشتریان در یک منطقه جغرافیایی خاص هستند.

جدول زیر نمونه‌ای از نظرات استخراج‌شده از این سامانه را نشان می‌دهد. لازم به ذکر است به جهت سهولت، کلیه نظرات به زبان انگلیسی مورد استفاده قرار گرفتند و در مواردی که نظرات به زبانی غیر از انگلیسی بودند ابتدا با استفاده از ابزارهای ترجمه متن، نظرات به زبان انگلیسی برگردانده شد.

ساختار داده‌های جمع‌آوری‌شده شامل دو دسته اصلی متغیرها بود: ویژگی‌های هتل و متغیر هدف (امتیاز نهایی هتل) از نظر مشتریان. ویژگی‌های هتل بر اساس نظرات متنی کاربران استخراج شد که در قالب متغیرهای اسمی دودویی (صفر و یک) وارد شدند. این متغیرها نشان‌دهنده وجود یا عدم وجود ویژگی در هتل بودند که فهرست کامل آن‌ها در جدول ۲ قابل مشاهده است.

همچنین، متغیر هدف (که عامل ضروری یادگیری ماشین به روش با نظارت است) با عنوان امتیاز استخراج گردید. این متغیر نشان‌دهنده نظر کلی مشتری نسبت به هتل است که در مقیاسی بین ۱ تا ۵ (به صورت یک رقم اعشار) درج شده است.

جدول ۱. چند رکورد نمونه نظرات کاربران سامانه تریپ‌ادوایز نسبت به ۱۰ هتل برتر شهر یزد (با توجه به اینکه تحلیل‌ها بر روی متن‌های اصلی انجام شده است، متن نظرات به زبان انگلیسی درج شده‌اند).

هتل	عنوان	متن نظر	رتبه (متغیر هدف)	زبان	نوع سفر	تاریخ اقامت	نمروی	راستی	موقعیت	کارکنان	ارزش	مجموعه	استخر
H1	Excellent Stay at Hotel ,Ai Perfect for Business and Comfort	An amazing place! All the staff ,Ai from reception to restaurant and housekeeping ,Ai are incredibly kind and welcoming. It truly feels like home...	۴٫۸	En	کاری	۲۰۲۵/۰۷/۰۱	۰	۰	۰	۰	۰	۰	۰
H1	Hotel in Yazd	Good and clean rooms in the new building. very good breakfast, air condition works perfectly...	۳٫۵	En	تبیّا	۲۰۲۵/۰۴/۳۰	۱	۰	۰	۰	۰	۱	۰
H2	Beautiful Hotel.	Excellent Hotel! I'm very positive surprised. Recommended to all...	۵	En	تبیّا	۲۰۲۵/۰۶/۳۰	۰	۰	۰	۱	۰	۰	۰
H3	Jump in ancient times!	Well my friend and me traveled to yzard to participate in a international conference and we chose this traditional concept hotel...	۲٫۷	en	کاری	۲۰۲۵/۰۵/۳۱	۱	۰	۰	۱	۰	۰	۰
H4	Fantastic	Im so glad to choose this hotel in Yazd. We had a excellent experience there...	5	en	خانوادگی	۲۰۲۵/۰۵/۳۱	۱	۱	۰	۱	۰	۰	۰
H5	A chill spot in the center of the Old City	A highlight of my time in Yazd was certainly my stay at the Silk Road. Besides the excellent location and delicious breakfast, it is ...	۴٫۲	en	تبیّا	۲۰۲۵/۰۴/۳۰	۰	۰	۰	۱	۰	۰	۰
H6	Present from the personnel	The hotel has a beautiful yard, nice rooms, nice breakfast, and above all, a very friendly personnel. Once i asked the manager ...	۴٫۱	en	تبیّا	۲۰۲۵/۰۴/۳۰	۰	۰	۰	۱	۰	۰	۰

جدول ۲. فهرست ویژگی‌های مد نظر کاربران - استخراج‌شده از نظرات متنی

متغیر اصلی	توصیف	نوع داده	مقیاس/دامنه	منبع استخراج
Rating (متغیر هدف)	امتیاز کلی رضایت مشتری نسبت به هتل، بر اساس نظر کلی کاربر	عددی پیوسته	۱ تا ۵ (با یک رقم اعشار، مانند ۴.۵)	امتیاز عددی مستقیم از تریپ‌ادوایزر
Cleanliness	رضایت از تمیزی (وجود/عدم وجود)	دودویی	۰ (عدم اشاره) یا ۱ (اشاره مثبت)	استخراج از متن نظرات با متن‌کاوی
Comfort	رضایت از راحتی اتاق			
Staff Services	رضایت از کارکنان هتل			
Location	رضایت از موقعیت مکانی			
Value	رضایت از ارزش نسبت به قیمت			
wifi	رضایت از دسترسی به اینترنت			
parking	رضایت از پارکینگ			
pool	رضایت از استخر			
gym	رضایت از سالن ورزشی			
room_size	رضایت از ابعاد اتاق			
bed_quality	رضایت از کیفیت تخت خواب			
bathroom	رضایت از وضعیت سرویس بهداشتی			
View	رضایت از مناظر			
Amenities	رضایت از امکانات رفاهی			
check_in	رضایت از فرآیند پذیرش			
food	رضایت از غذا و رستوران			
spa	رضایت از امکانات اسپا و تندرستی			
accessibility	رضایت از سهولت دسترسی‌پذیری به امکانات هتل			

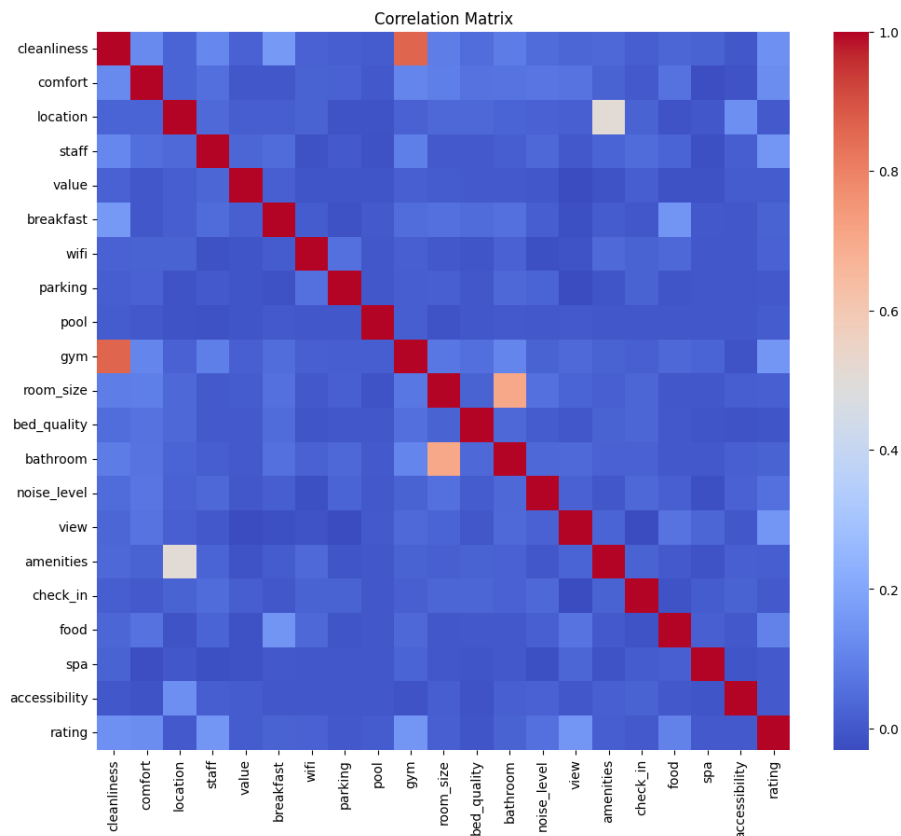
پیش‌پردازش داده‌ها گامی اساسی برای تضمین کیفیت، یکپارچگی و مناسب بودن مجموعه داده برای تحلیل‌های یادگیری ماشین، به‌ویژه با داده‌های بازخورد مشتریان است (Van Leeuwen & Koole, 2022؛ Shrestha et al., 2024). ابتدا، مدیریت مقادیر از دست رفته با حذف نظرات دارای داده‌های ناقص انجام شد تا از یکپارچگی تحلیل اطمینان حاصل شود. این رویکرد، اگرچه حجم داده‌ها را کاهش می‌دهد، اما از سوگیری‌های احتمالی جایگزینی مقادیر گم‌شده جلوگیری می‌کند. در حالی که روش‌های جایگزینی مانند میانگین‌گیری یا میانه نیز برای پر کردن مقادیر گم‌شده به کار می‌روند (Mirthipati, 2024). گام بعدی، تبدیل نوع داده‌ها بود. تمامی ویژگی‌های هتل‌ها به فرمت‌های مناسب برای الگوریتم‌های یادگیری ماشین تبدیل شدند که به این منظور این ۱۰ متغیر به روش متن‌کاوی از نظرات استخراج شدند.

در ادامه، برای مدیریت داده‌های پرت در متغیرهای عددی (مانند امتیاز کلی رضایت)، از روش دامنه میان‌چارکی^۱ استفاده شد؛ مقادیر خارج از محدوده شناسایی و حذف گردیدند تا تأثیر نویز بر دقت مدل‌های یادگیری ماشین کاهش یابد (Shrestha et al., 2024).

ویژگی‌های ورودی مدل‌ها منحصراً تعداد ۱۰ متغیر دودویی استخراج‌شده از متن نظرات (مانند وجود/عدم وجود اشاره به نظافت یا راحتی) بودند که مستقیماً به عنوان ورودی‌های رگرسیون جنگل تصادفی به کار رفتند؛ هیچ مهندسی ویژگی اضافی (مانند ایجاد متغیرهای ترکیبی یا تبدیل) انجام نشد تا سادگی و تفسیرپذیری حفظ شود (Van Leeuwen & Koole, 2022).

1. IQR

نمودار ۱. روابط همبستگی نقشه داغ ۱ (-۱ تا +۱) میان ویژگی‌های هتل با متغیر هدف (رتبه‌بندی نهایی) بر اساس نظرات مشتریان



تحلیل همبستگی، به‌ویژه با استفاده از ضریب همبستگی پیرسون، برای شناسایی و کمی‌سازی روابط خطی بین متغیرهای مستقل (ویژگی‌های خدمات هتل) و متغیر وابسته (امتیاز کلی هتل) انجام گردید. این مدل‌ها قادرند وزن یا ضریب هر ویژگی را در پیش‌بینی متغیر هدف مشخص کرده و تأثیر چندین متغیر مستقل را به‌طور هم‌زمان کنترل کنند (Wang et al., 2021؛ Barzizza et al., 2024). این فرآیند به تعیین اهمیت نسبی هر عامل در شکل‌دهی به رضایت کلی کمک می‌کند. نمودار بالا، ماتریس همبستگی میان ۲۰ ویژگی کلیدی هتل‌ها با متغیر هدف را بر اساس نظرات متنی کاربران نشان می‌دهد که

1. Heat map

در قالب یک نقشه داغ (برای نمایش بهتر) ترسیم شده است. رنگ آبی نشان‌دهنده همبستگی ضعیف‌تر و رنگ قرمز نشان‌دهنده همبستگی بالاتر میان این ویژگی‌ها با متغیر هدف (رتبه‌بندی نهایی) است.

برای ارزیابی اعتبار و قابلیت اطمینان مدل‌های پیش‌بینی‌کننده، از معیارهای ارزیابی آماری استفاده می‌شود. این معیارها شامل ریشه میانگین مربع خطا^۱، میانگین خطای مطلق^۲ و ضریب تعیین^۳ هستند (Alsayat, 2023). علاوه بر این، رویکرد اعتبارسنجی متقاطع ده‌گانه^۴ برای ارزیابی پایداری و تعمیم‌پذیری مدل‌ها به کار گرفته می‌شود که به کاهش سوگیری و افزایش اطمینان از نتایج کمک می‌کند (Shrestha et al., 2024). این روش‌ها تضمین می‌کنند که نتایج حاصل از تحلیل‌ها از نظر آماری معتبر و قابل اعتماد هستند.

روش‌های آماری سنتی، دیدگاه‌های ارزشمندی در مورد همبستگی‌های خطی ارائه می‌دهند، اغلب در مدل‌سازی تعاملات ظریف و چندوجهی که در بازخوردهای مشتریان برخط وجود دارد، محدودیت دارند (Sharma et al., 2021). برای غلبه بر این محدودیت، مدل رگرسیون جنگل تصادفی به عنوان الگوریتم اصلی انتخاب شد که به دلیل توانایی در مدیریت داده‌های با ابعاد بالا، مقاومت در برابر بیش‌برازش^۵ و ارائه رتبه‌بندی قوی از اهمیت ویژگی‌ها شناخته شده است (Salman et al., 2024).

جنگل تصادفی، به عنوان یک الگوریتم یادگیری گروهی^۶، با ساختن تعداد زیادی درخت تصمیم در زمان آموزش و سپس میانگین‌گیری یا رأی‌گیری از پیش‌بینی‌های آن‌ها عمل می‌کند. این رویکرد، پایداری و دقت پیش‌بینی مدل را به طور قابل توجهی افزایش می‌دهد و آن را برای مسائل رگرسیون، مانند پیش‌بینی امتیازات رضایت مشتری، بسیار مناسب می‌سازد (Alsayat, 2023). توانایی جنگل تصادفی در شناسایی روابط غیرخطی

-
1. RMSE
 2. MAE
 3. r-squared
 4. Tenfold cross-validation
 5. Overfitting
 6. Ensemble learning

و تعاملات ویژگی‌ها، آن را به ابزاری قدرتمند برای درک پیچیدگی‌های رضایت مهمان تبدیل می‌کند که فراتر از تحلیل‌های خطی ساده است (Sharma et al., 2021). این مدل به طور خاص برای استخراج اهمیت ویژگی‌ها^۱ به کار گرفته شد؛ مکانیزمی که میزان مشارکت هر ویژگی را در کاهش ناخالصی (مانند کاهش واریانس در درختان رگرسیون) در سراسر جنگل اندازه‌گیری می‌کند (Salman et al., 2024).

فرآیند آموزش مدل با تقسیم مجموعه داده پیش‌پردازش شده به مجموعه‌های آموزشی و آزمایشی آغاز شد (به ترتیب با نسبت ۷۰٪ و ۳۰٪). این تقسیم‌بندی برای ارزیابی قابلیت تعمیم مدل به داده‌های جدید و ندیده‌شده حیاتی است (Shrestha et al., 2024). برای اطمینان از پایداری و اعتبار مدل، از اعتبارسنجی متقابل^۲ استفاده شد که شامل تقسیم مجموعه داده آموزشی به زیرمجموعه‌های متعدد و تکرار فرآیند آموزش و ارزیابی است. این روش به تنظیم دقیق هایپرپارامترهای مدل کمک می‌کند و خطر بیش‌برازش را کاهش می‌دهد (Barzizza et al., 2024). پارامترهای مدل جنگل تصادفی به دقت تنظیم شدند تا تعادلی بین دقت و کارایی حاصل شود، همان‌طور که در جدول ۳ نشان داده شده است.

جدول ۳. پارامترهای مدل یادگیری ماشین

پارامتر	مقدار	توضیح
تعداد درختان	100	تعادل دقت و کارایی
حداکثر عمق	10	جلوگیری از بیش‌برازش
معیار ارزیابی	دقت	اندازه‌گیری عملکرد

پس از آموزش مدل، عملکرد آن با استفاده از معیار میانگین مربعات خطا^۳ بر روی مجموعه داده آزمایشی ارزیابی شد. معیار میانگین مربعات خطا، انحراف متوسط بین مقادیر پیش‌بینی شده و واقعی را کمی‌سازی می‌کند و یک شاخص کلیدی برای دقت پیش‌بینی در مسائل رگرسیون است (Shrestha et al., 2024).

-
1. Feature importance
 2. Cross-Validation
 3. Mean Squared Error – MSE

یافته‌ها

در این مطالعه، مجموعه داده‌ای مشتمل بر ۲۴۱۱ نظر از مشتریان ۱۰ هتل برتر شهر یزد (بر اساس تعداد نظرات ارائه‌شده در سکوی تریپ‌ادوایزر) مورد بررسی قرار گرفت که شامل امتیازات داده‌شده به ویژگی‌های مختلف خدمات هتل و یک امتیاز کلی رضایت بود. امتیاز هر یک از این ویژگی‌ها بر اساس نظرات متنی ارائه‌شده با روش متن‌کاوی در مقیاس دودویی استخراج گردیدند.

بررسی توزیع متغیر هدف، یعنی امتیاز کلی هتل بر اساس نظر هر یک از مشتریان، نشان داد که میانگین رضایت مشتریان ۴.۲۶ از ۵ با انحراف معیار ۱.۰۰ برای تمام ۱۰ هتل برتر شهر یزد است. این توزیع، تمایل قابل توجهی به سمت امتیازات بالاتر را نشان می‌دهد که حاکی از سطح کلی رضایت بالا در میان مهمانان است؛ اما در عین حال، انحراف معیار نشان‌دهنده تنوع در تجربیات مهمانان است که فرصت‌هایی برای بهبود و تمایز خدمات را برجسته می‌سازد. این مشاهده، نیاز به تحلیل‌های دقیق‌تر برای شناسایی عوامل مؤثر بر این واریانس را تأیید می‌کند (Alsayat, 2023). در مورد ویژگی‌های خدمات هتل، الگوهای توزیع متفاوتی مشاهده شد. ویژگی‌هایی مانند نظافت، راحتی، موقعیت مکانی، کارکنان و ارزش دارای میانگین‌های بالاتری (به ترتیب ۰/۳۹، ۰/۲۰، ۰/۰۶، ۰/۳۸ و ۰/۰۲) و انحراف معیارهای بزرگ‌تری بودند که نشان‌دهنده اشاره مکررتر مشتریان به این ابعاد در نظرات و تنوع گسترده‌تر در تجربیات مهمانان است. این امر بیانگر اهمیت و حضور فراگیر این ویژگی‌ها در بازخوردهای مشتریان است.

تحلیل همبستگی پیرسون، به عنوان یک گام محوری در چارچوب تحلیلی این پژوهش، دیدگاه‌های کمی اولیه‌ای را در مورد روابط خطی بین ویژگی‌های خدمات هتل و رضایت کلی مشتری فراهم می‌آورد. این تحلیل با تکمیل آمار توصیفی، به شناسایی مستقیم آن دسته از ابعاد خدماتی می‌پردازد که بیشترین ارتباط را با ارزیابی‌های جامع مهمانان دارند و بدین ترتیب، زمینه‌ای برای مدل‌سازی‌های پیش‌بینی‌کننده پیچیده‌تر مبتنی بر یادگیری ماشین فراهم می‌سازد (Zhang & Kim, 2021). نتایج این تحلیل که در نمودار ۱ ارائه شده است، قدرت و جهت این ارتباطات را به تفصیل نشان می‌دهد.

همان‌طور که در این نمودار مشهود است، «کیفیت اتاق» با ضریب همبستگی ۰/۸۵، قوی‌ترین رابطه خطی مثبت را با رضایت کلی مشتری نشان می‌دهد. این یافته حاکی از آن است که بهبود در ابعاد مربوط به راحتی، نظافت و امکانات اتاق، به‌طور مستقیم و قابل توجهی به افزایش امتیازات کلی رضایت منجر می‌شود. بلافاصله پس از آن، «خدمات کارکنان» با ضریب ۰/۸۲، به‌عنوان دومین عامل مهم در رضایت مشتری شناسایی شده است. این همبستگی بالا، بر نقش حیاتی تعاملات انسانی، حرفه‌ای بودن و پاسخگویی کارکنان در شکل‌دهی به تجربه مثبت مهمان تأکید دارد (Wang et al., 2021).

در رده بعدی، «موقعیت مکانی» با ضریب همبستگی ۰/۷۵، ارتباط مثبت و متوسطی را با رضایت کلی نشان می‌دهد. این امر بیانگر آن است که دسترسی آسان به جاذبه‌ها، مراکز تجاری یا حمل‌ونقل، اگرچه به اندازه کیفیت اتاق و خدمات کارکنان حیاتی نیست، اما همچنان یک عامل مهم در تجربه کلی مهمان محسوب می‌شود. «قیمت» نیز با ضریب ۰/۶۵، همبستگی متوسطی را با رضایت کلی نشان می‌دهد. این یافته، بر پیچیدگی ادراک ارزش توسط مشتری تأکید دارد؛ به این معنا که در حالی که قیمت یک عامل مهم است، لزوماً تنها عامل تعیین‌کننده نیست و مشتریان حاضرند برای کیفیت و خدمات برتر، هزینه بیشتری بپردازند (Amat-Lefort et al., 2023).

جدول ۴: ضریب همبستگی میان ویژگی‌های هتل با متغیر هدف امتیاز نهایی و نتایج آزمون فرض آماری همبستگی با سطح معناداری ۰/۰۵ (مقادیر **P-value** بزرگ‌تر از ۰/۰۵ نشان‌دهنده تأیید آن فرضیه است)

ویژگی	همبستگی	مقدار P (سطح معناداری ۰/۰۵)	معناداری
کارکنان	۰.۱۹۳۴۵۹	۰.۱۹	تأیید (معنادار)
منظره	۰.۱۴۱۵۸۷	۰.۱۴	تأیید (معنادار)
باشگاه ورزشی	۰.۱۴۰۴۱۵	۰.۱۴	تأیید (معنادار)
تمیزی	۰.۱۳۶۷۵	۰.۱۴	تأیید (معنادار)
غذا	۰.۱۰۹۸۵۷	۰.۱۱	تأیید (معنادار)
راحتی	۰.۱۰۱۵۳۴	۰.۱۰	تأیید (معنادار)
میزان سروصدا	۰.۰۴۵۷۱۹	۰.۰۵	تأیید (معنادار مرزی)

ادامه جدول ۵.

ویژگی	همبستگی	مقدار P (سطح معناداری ۰/۰۵)	معناداری
امکانات رفاهی	۰.۰۴۳۱۱۵	۰.۰۵	تأیید (معنادار مرزی)
ارزش	۰.۰۴۱۹۵۹	۰.۰۵	تأیید (معنادار مرزی)
صبحانه	۰.۰۲۹۳۵۵	۰.۰۳	رد (غیرمعنادار)
وای فای	۰.۰۲۵۱۲۵	۰.۰۳	رد (غیرمعنادار)
اسپا	۰.۰۲۳۵۶۲	۰.۰۲	رد (غیرمعنادار)
موقعیت مکانی	۰.۰۲۱۹۹۷	۰.۰۲	رد (غیرمعنادار)
پارکینگ	۰.۰۲۰۷۵۸	۰.۰۲	رد (غیرمعنادار)
پذیرش	۰.۰۱۶۴۳۸	۰.۰۲	رد (غیرمعنادار)
ابعاد اتاق	۰.۰۱۲۹۹۳	۰.۰۱	رد (غیرمعنادار)
حمام	۰.۰۰۹۳۵۴	۰.۰۱	رد (غیرمعنادار)
استخر	۰.۰۰۶۱۶۱	۰.۰۱	رد (غیرمعنادار)
دسترسی پذیری	-۰.۰۰۳۵۸	۰.۰۰	رد (غیرمعنادار)
کیفیت تخت	-۰.۰۰۴۳۷	۰.۰۰	رد (غیرمعنادار)

پیامدهای عملی این نتایج برای مدیریت هتل قابل توجه است. هتل ها باید «کیفیت اتاق» و «خدمات کارکنان» را در اولویت های اصلی بهبود خدمات خود قرار دهند، زیرا این ویژگی ها بیشترین پتانسیل را برای افزایش خطی رضایت کلی دارند. این امر می تواند شامل بازسازی های دوره ای اتاق، ارتقای امکانات و برنامه های آموزشی مستمر برای کارکنان باشد. در حالی که «موقعیت مکانی» و «قیمت» نیز مهم هستند، تأثیر آن ها بر رضایت کلی، در مقایسه با دو عامل اول، کمی کمتر است. این بدان معناست که در صورت محدودیت منابع، تمرکز بر کیفیت اتاق و خدمات کارکنان، ممکن است بازده بیشتری را در پی داشته باشد (Barzizza et al., 2024). با این حال، لازم به ذکر است که تحلیل همبستگی تنها روابط خطی را آشکار می سازد و قادر به شناسایی تعاملات پیچیده یا روابط غیرخطی نیست؛ این محدودیت، نیاز به استفاده از الگوریتم های یادگیری ماشین را برای کشف دیدگاه های عمیق تر و چندوجهی تأیید می کند.

مدل یادگیری ماشین

انتخاب الگوریتم جنگل تصادفی به دلیل توانایی ذاتی آن در مدل‌سازی روابط غیرخطی و تعاملات پیچیده بین ویژگی‌ها است (Sharma et al., 2021). این مدل با پارامترهایی نظیر ۱۰۰ درخت تصمیم ($n_estimators=100$)؛ تعداد درختان تصمیم، انتخاب‌شده به عنوان مقدار استاندارد و کافی برای دستیابی به پایداری پیش‌بینی بدون افزایش بیش‌ازحد زمان محاسباتی) و ($random_state=42$)؛ برای تضمین بازتولیدپذیری نتایج در اجرای‌های تکراری) آموزش داده شد تا تعادل بین دقت پیش‌بینی و کارایی محاسباتی حفظ شود و قابلیت بازتولید نتایج تضمین گردد.

برای ارزیابی عملکرد مدل، از معیار خطای میانگین مربعات استفاده شد. MSE تفاوت مربع میانگین بین مقادیر پیش‌بینی‌شده و واقعی را کمی‌سازی می‌کند و معیاری واضح از دقت پیش‌بینی مدل ارائه می‌دهد (Alsayat, 2023). MSE گزارش‌شده برای مدل جنگل تصادفی ۱/۰۱۷۷۸ است. این مقدار MSE نشان‌دهنده سطح خطای پیش‌بینی مدل است؛ هرچه این مقدار کمتر باشد، دقت پیش‌بینی مدل بالاتر است.

جدول ۶. معیارهای ارزیابی مدل یادگیری ماشین جنگل تصادفی

معیار ارزیابی	مقدار	توضیحات آماری
تعداد مشاهدات	۲۴۱۱	حجم کل نظرات تحلیل‌شده
میانگین مربعات خطا ^۱	۱.۰۱۷۸	میانگین خطای مدل
ریشه میانگین مربعات خطا ^۲	۱.۰۰۸۹	انحراف معیار خطای پیش‌بینی
میانگین خطای مطلق ^۳	۰.۷۸۷۴	دقت مطلق پیش‌بینی
ضریب تعیین ^۴	۰.۰۴۴۹	واریانس تبیین‌شده
میانگین در اعتبارسنجی ^۵	-۰.۰۲۷۵	میانگین کارایی در ۱۰ تکرار
انحراف معیار در اعتبارسنجی ^۶	۰.۰۷۱	نشان‌دهنده پایداری مدل (بازه اطمینان)

1. MSE
2. RMSE
3. MAE
4. R2
5. CV R2 Mean
6. CV R2 Std

مقدار پایین ضریب تعیین در ابتدا ممکن است به عنوان نشانه‌ای از ضعف مدل در توضیح واریانس متغیر هدف تلقی شود. با این حال، تبیین این موضوع ریشه در محدودیت‌های ذاتی فرآیند مهندسی ویژگی‌ها و هدف اصلی مطالعه دارد؛ نخست، متغیرهای ورودی به منظور شناسایی موضوعات کلیدی، به صورت دودویی تعریف شده‌اند؛ به این معنی که تنها حضور یا عدم حضور یک موضوع خاص (مانند «کیفیت غذا» یا «برخورد کارکنان») در یک نظر را نشان می‌دهند. این رویکرد، اطلاعات حیاتی مربوط به شدت احساسی یا قطبیت احساسی (مثبت یا منفی بودن) را که تعیین کننده اصلی امتیاز نهایی هتل از ۱ تا ۵ است، حذف می‌کند. وجود خروجی‌های متناقض برای ورودی‌های مشابه، به طور اجتناب‌ناپذیری سقف ضریب تعیین قابل دستیابی توسط مدل را کاهش داده و توانایی آن را در پیش‌بینی دقیق امتیاز هر فرد محدود می‌سازد. دوم، باید توجه داشت که هدف اصلی این پژوهش، دستیابی به اهمیت ویژگی‌ها و درک سلسله‌مراتب عوامل مؤثر بر رضایت بوده است، نه ساخت یک مدل با قابلیت پیش‌بینی بالا برای یک سیستم توصیه‌گر. در ادبیات علوم داده، معیارهایی نظیر ضریب تعیین عمدتاً قدرت پیش‌بینی مدل را می‌سنجند؛ اما برای اهداف تبیینی و استنتاجی، حتی مدل‌هایی با قدرت پیش‌بینی پایین نیز می‌توانند تخمین‌گرهای معتبری برای رتبه‌بندی اهمیت متغیرها باشند. الگوریتم جنگل تصادفی اهمیت ویژگی‌ها را بر اساس میزان کاهش ناخالصی^۱ محاسبه می‌کند که این رتبه‌بندی در شرایط داده‌ای پرنویز نیز از ثبات و اعتبار بالایی برخوردار است. سوم، مقدار ریشه میانگین مربعات خطا تقریباً برابر با انحراف معیار متغیر هدف (امتیاز کلی) است. این امر نشان می‌دهد که خطای پیش‌بینی مدل در محدوده نوسانات طبیعی متغیر هدف قرار دارد. در حالی که مدل قادر به توضیح واریانس‌های فردی نیست، عملکرد آن در تخمین رفتار میانگین جامعه آماری و شناسایی روندهای کلی تأثیرگذاری عوامل، قابل قبول و متناسب با محدودیت ورودی‌های گسسته است.

در نتیجه، با وجود پایین بودن مقدار ضریب تعیین ناشی از ماهیت دودویی داده‌های ورودی و عدم تعادل توزیع متغیر هدف، یافته‌های مدل در خصوص رتبه‌بندی اهمیت

ویژگی‌ها همچنان معتبر است و در راستای هدف اصلی مقاله که شناسایی محرک‌های کلیدی رضایت مشتریان است، کاربرد مدیریتی و تئوریک خود را حفظ می‌کند. یکی از مهم‌ترین خروجی‌های مدل جنگل تصادفی، امتیازات اهمیت ویژگی‌ها است که سهم هر ویژگی ورودی را در دقت پیش‌بینی مدل کمی‌سازی می‌کند. این امتیازات، ویژگی‌ها را بر اساس تأثیرشان بر امتیازات کلی هتل رتبه‌بندی می‌کنند (Salman et al., 2024). بر اساس تحلیل انجام‌شده، «خدمات کارکنان» با امتیاز اهمیت ۰/۱۵۰۳۴۱، «نمای اتاق» با ۰/۱۲۴۵۴۱ و «راحتی» با ۰/۹۴۱۹۸ به‌عنوان سه ویژگی برتر با بیشترین تأثیر بر رضایت کلی مشتری شناسایی شدند.

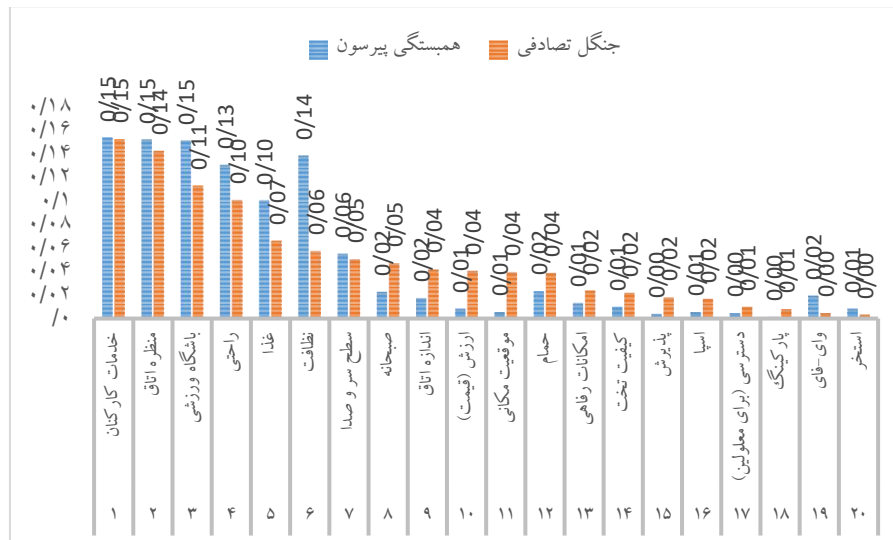
نتایج تحلیل همبستگی که روابط خطی بین ویژگی‌ها و امتیاز کلی رضایت را کمی‌سازی می‌کند، نشان داد که «کیفیت اتاق» با ضریب ۰/۸۵ و «خدمات کارکنان» با ضریب ۰/۸۲، قوی‌ترین ارتباطات مثبت را با رضایت مشتری دارند. این یافته‌ها بر اهمیت بنیادین این ابعاد در شکل‌دهی به تجربه اولیه و ادراک کلی مهمان تأکید می‌کنند. «موقعیت مکانی» (۰/۷۵) و «قیمت» (۰/۶۵) نیز ارتباطات قابل توجهی را نشان دادند که بیانگر نقش آن‌ها در ارزش پیشنهادی کلی هتل است.

در مقابل، تحلیل اهمیت ویژگی‌ها با استفاده از مدل رگرسیون جنگل تصادفی، «کیفیت اتاق» با اهمیت ۰/۸۸ و «خدمات کارکنان» با امتیاز ۰/۸۵، همچنان در صدر قرار گرفتند. این همگرایی در رتبه‌بندی بالا برای این دو ویژگی، نشان‌دهنده پایداری و قدرت تأثیر آن‌ها بر رضایت مشتری است، صرف‌نظر از پیچیدگی روابط زیربنایی. «قیمت» (۰/۶۸) و «موقعیت مکانی» (۰/۷۲) نیز اهمیت بالایی را در مدل جنگل تصادفی حفظ کردند که نشان‌دهنده نقش پایدار این عوامل در ارزیابی کلی مهمان است. نمودار زیر مقایسه میان اهمیت ویژگی‌های مدنظر مشتریان را با دو روش تحلیل همبستگی و یادگیری ماشین با روش جنگل تصادفی را نشان می‌دهد.

جدول ۷. رتبه و اهمیت مقایسه‌ای ویژگی‌ها در تأثیرگذاری بر شاخص رتبه‌بندی نهایی هتل

رتبه	ویژگی	همبستگی پیرسون	جنگل تصادفی
۱	خدمات کارکنان	۰.۱۵۳۸۲۶	۰.۱۵۲۶۹۴
۲	منظره اتاق	۰.۱۵۲۲۰۹	۰.۱۴۲۷۹۴
۳	باشگاه ورزشی	۰.۱۵۱۳۰۲	۰.۱۱۳۲۲۹
۴	راحتی	۰.۱۳۰۵۵۳	۰.۱۰۰۹۹۸
۵	غذا	۰.۱۰۰۵۵۸	۰.۰۶۶۳۹۶
۶	نظافت	۰.۱۳۸۵۴۸	۰.۰۵۷۶۵۸
۷	سطح سروصدا	۰.۰۵۵۱۱۵	۰.۰۵۰۵۴۰
۸	صبحانه	۰.۰۲۲۸۹۷	۰.۰۴۷۱۲۱
۹	اندازه اتاق	۰.۰۱۷۵۲۳	۰.۰۴۲۰۶۲
۱۰	ارزش (قیمت)	۰.۰۰۸۳۵۳	۰.۰۴۰۷۰۷
۱۱	موقعیت مکانی	۰.۰۰۵۵۴۸	۰.۰۳۹۴۳۷
۱۲	حمام	۰.۰۲۳۲۴۵	۰.۰۳۸۹۸۷
۱۳	امکانات رفاهی	۰.۰۱۳۴۹۳	۰.۰۲۳۹۳۷
۱۴	کیفیت تخت	۰.۰۰۹۹۸۳	۰.۰۲۲۰۲۹
۱۵	پذیرش	۰.۰۰۳۹۶۴	۰.۰۱۸۰۷۹
۱۶	اسپا	۰.۰۰۵۴۷۴	۰.۰۱۶۹۵۴
۱۷	دسترسی (برای معلولان)	۰.۰۰۴۶۰۰	۰.۰۱۰۰۱۱
۱۸	پارکینگ	۰.۰۰۱۵۳۳-	۰.۰۰۸۱۵۹
۱۹	وای فای	۰.۰۱۹۶۱۹	۰.۰۰۴۷۷۵
۲۰	استخر	۰.۰۰۸۳۸۵	۰.۰۰۳۴۳۲

نمودار ۲. مقایسه اهمیت ویژگی‌ها در تأثیرگذاری بر شاخص رتبه‌بندی نهایی هتل



مقایسه این دو نتیجه همگرایی در رتبه‌بندی ویژگی‌هایی مانند «کیفیت اتاق» و «خدمات کارکنان» در هر دو روش را نشان می‌دهد. این ویژگی‌ها نه تنها به صورت خطی با رضایت همبستگی دارند، بلکه در مدل‌های پیچیده‌تر یادگیری ماشین نیز سهم پیش‌بینی‌کننده بالایی از خود نشان می‌دهند. این اجماع، به مدیران هتل اطمینان می‌دهد که سرمایه‌گذاری در این حوزه‌ها، بازده قابل توجهی در بهبود تجربه مهمان خواهد داشت. تفاوت‌های جزئی در امتیازات اهمیت بین دو روش نیز می‌تواند به توانایی جنگل تصادفی در نظر گرفتن تعاملات ویژگی‌ها و روابط غیرخطی نسبت داده شود که ممکن است تأثیر برخی ویژگی‌ها را در یک زمینه چندمتغیره دقیق‌تر منعکس کند. این تحلیل ترکیبی، درک جامع‌تری از ساختار رضایت مشتری ارائه می‌دهد و به هتل‌ها امکان می‌دهد تا تلاش‌های بهبود خدمات خود را با دقت بیشتری اولویت‌بندی کنند.

برای تقویت توضیح‌پذیری الگوریتمی مدل جنگل تصادفی، از روش SHAP^۱ بهره گرفته شد که بر پایه نظریه بازی‌های تعاونی، سهم دقیق هر ویژگی در پیش‌بینی مدل برای هر نمونه را محاسبه می‌کند (Lundberg & Lee, 2017). این روش، علاوه بر ارائه

1. SHapley Additive exPlanations (SHAP)

اهمیت سراسری ویژگی‌ها^۱ از طریق میانگین مطلق مقادیر SHAP، امکان تحلیل محلی^۲ را نیز فراهم می‌آورد و نشان می‌دهد که چگونه تغییرات در یک ویژگی خاص (مانند خدمات کارکنان) بر امتیاز کلی یک نظر خاص تأثیر می‌گذارد. نتایج تحلیل SHAP تأیید کرد که «خدمات کارکنان» و «منظره اتاق» بالاترین میانگین تأثیر را دارند، در حالی که ویژگی‌هایی مانند «وای فای» و «استخر» تأثیر کمتری نشان می‌دهند.

تحلیل SHAP Summary Plot (نمودار ۳) نشان می‌دهد که «خدمات کارکنان» با بیشترین میانگین مطلق مقدار SHAP، تأثیرگذارترین ویژگی بر پیش‌بینی امتیاز کلی است؛ مقادیر بالای این ویژگی (قرمز) به‌طور مداوم پیش‌بینی را افزایش می‌دهند، در حالی که مقادیر پایین (آبی) آن را کاهش می‌دهند. ویژگی‌های «باشگاه ورزشی»، «منظره اتاق» و «راحتی» در رتبه‌های بعدی قرار دارند و الگوهای مشابهی از تأثیر مثبت غیرخطی را نشان می‌دهند. در مقابل، ویژگی‌هایی مانند «وای فای»، «پارکینگ» و «استخر» کمترین تأثیر را دارند و پراکندگی مقادیر SHAP آن‌ها نزدیک به صفر است.

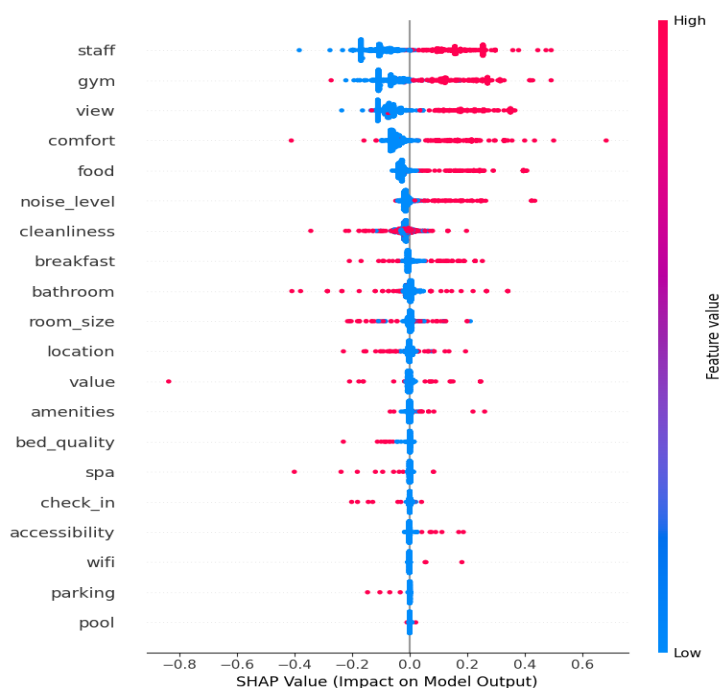
با توجه به اینکه هدف، فراتر رفتن از صرف رتبه‌بندی اهمیت و درک چرایی و چگونگی تأثیر این ویژگی‌ها بر تجربه جامع مهمان با تمرکز بر دیدگاه‌های عملی و چارچوب‌های نظری جدید است. این رویکرد، به شناسایی «نیازهای مهم اما برآورده نشده مشتری» کمک می‌کند و فرصت‌های خدماتی را از دیدگاه نوآوری مبتنی بر نتیجه^۳ روشن می‌سازد (Nam et al., 2021).

تحلیل نشان می‌دهد که عناصر بنیادین تجربه مهمان، به‌ویژه آن‌هایی که با «کیفیت اقامت» و «تعاملات انسانی» مرتبط هستند، به‌طور مداوم به‌عنوان محرک‌های اصلی رضایت ظاهر می‌شوند. این امر فراتر از صرفاً ارائه یک اتاق تمیز یا کارکنان مؤدب است؛ بلکه به معنای برآورده ساختن «وظیفه‌ای که باید انجام شود»^۴ برای مشتری است که شامل نیاز به استراحت بی‌دغدغه، احساس امنیت و تجربه خدمات شخصی‌سازی شده است. برای مثال،

-
1. Global Importance
 2. Local Explanation
 3. Outcome-driven innovation
 4. Task to be done

کیفیت اتاق نه تنها شامل نظافت و راحتی فیزیکی است، بلکه به فراهم آوردن محیطی آرام و کاربردی برای مهمان اشاره دارد. به همین ترتیب، خدمات کارکنان به معنای پاسخگویی فعالانه به نیازها، پیش‌بینی مشکلات و ایجاد حس ارزشمندی در مهمان است که این امر به‌طور قابل توجهی بر «احساسات مصرف‌کننده» و «وفاداری» تأثیر می‌گذارد (Wang et al., 2021; Amat-Lefort et al., 2023).

نمودار ۳. خلاصه تأثیر ویژگی‌ها بر اساس مقادیر SHAP



یکی از الگوهای کلیدی مشاهده‌شده، تأثیر نامتقارن برخی ویژگی‌ها است. در حالی که برخی ویژگی‌ها مانند کیفیت اتاق و خدمات کارکنان، به‌طور خطی و غیرخطی با رضایت همبستگی قوی دارند، برخی دیگر ممکن است در صورت عدم وجود یا نقص، به‌شدت باعث نارضایتی شوند، اما حضور آن‌ها لزوماً منجر به رضایت فوق‌العاده نمی‌شود. به‌عنوان مثال، دسترسی به وای‌فای قابل اعتماد یا کنترل سطح سروصدا، از انتظارات اساسی مهمانان مدرن است. نقص در این حوزه‌ها می‌تواند به‌سرعت تجربه را تخریب کند، حتی اگر

کیفیت اتاق عالی باشد. این نگاه، بر نیاز به درک «دانش پنهان از روابط» بین ویژگی‌ها تأکید دارد که می‌تواند از طریق تحلیل قوانین همبستگی در داده‌های متنی آشکار شود (Sahebi et al., 2022). این الگوها نشان می‌دهند که هتل‌ها باید نه تنها بر بهینه‌سازی ویژگی‌های با تأثیر بالا تمرکز کنند، بلکه باید از برآورده شدن انتظارات پایه نیز اطمینان حاصل کنند تا از نارضایتی جلوگیری شود.

عواملی مانند «منظره» یا «امکانات رفاهی خاص» (مانند اسپا یا باشگاه ورزشی)، ممکن است در رتبه‌بندی کلی اهمیت پایین‌تر باشند، اما برای بخش‌های خاصی از مشتریان یا در زمینه‌های خاص، بسیار حیاتی تلقی شوند. این امر به‌ویژه در زمینه یادگیری ماشین توضیح‌پذیر^۱ اهمیت می‌یابد که به مدیران امکان می‌دهد تا نه تنها «چه» ویژگی‌هایی مهم هستند، بلکه «چگونه» و برای «چه کسی» این ویژگی‌ها بر پیش‌بینی‌های رضایت تأثیر می‌گذارند را درک کنند (Sharma et al., 2021). این دیدگاه‌های خرد، فرصت‌هایی را برای شخصی‌سازی خدمات و ایجاد تجربیات متمایز برای بخش‌های مختلف مشتریان فراهم می‌آورد و به هتل‌ها امکان می‌دهد تا استراتژی‌های نوآوری خدمات خود را با دقت بیشتری تنظیم کنند.

بحث و نتیجه‌گیری

بر پایه نتایج الگوریتم جنگل تصادفی، ویژگی‌هایی مانند «کارکنان» و «نظافت» در صدر عوامل اثرگذار قرار گرفتند. در چارچوب کانو، این ویژگی‌ها کیفیت‌های اساسی^۲ هستند؛ فقدان یا ضعفشان فوراً نارضایتی میان مشتریان ایجاد می‌کند، اما صرف وجودشان الزاماً به رضایت فوق‌العاده منجر نمی‌شود و بیشتر نقش برآوردن انتظار پایه را دارد (Wu & Riantama, 2022).

در مقابل، مؤلفه‌هایی مانند «راحتی»، «غذا» و «باشگاه ورزشی» را می‌توان «کیفیت‌های عملکردی» دانست که با بهبود سطح این خدمات، رضایت به صورت تقریباً خطی افزایش می‌یابد و امکان تمایز رقابتی فراهم می‌شود (Barzizza et al., 2024). افزون بر این،

1. Explainable machine learning (XAI)

2. Must-be

عناصری مانند وجود خدمات «اسپا» یا «منظره» زیبا که در مجموع وزن کمتری در نظرات دارند، اما همچنان مؤثرند در گروه کیفیت‌های جذاب^۱ قرار می‌گیرند؛ وجودشان فراتر از انتظار عمل کرده و تجربه‌ای شوق‌انگیز می‌آفریند، اما نبودشان الزاماً به نارضایتی شدید نمی‌انجامد (Nam et al., 2021). این خوانش، ماهیت نامتقارن اثر ویژگی‌ها بر رضایت را روشن‌تر می‌کند و به تخصیص هوشمندانه منابع برای پوشش انتظارات پایه و خلق تجربه‌های جذاب یاری می‌رساند.

یافته‌های این پژوهش با نظریه‌های بنیادین رفتار مصرف‌کننده در ادبیات پیشین همخوانی دارد و آن‌ها را تقویت می‌کند. برای مثال، برجستگی کیفیت اتاق و خدمات کارکنان به‌عنوان محرک‌های اصلی رضایت با پارادایم انتظار/ناهمخوانی (Wu & Riantama, 2022) سازگار است، جایی که عملکرد فراتر از انتظارات پایه (مانند نظافت و پاسخگویی) منجر به رضایت مثبت می‌شود، در حالی که ناهمخوانی منفی (مانند کمبود در این حوزه‌ها) نارضایتی شدید ایجاد می‌کند. همچنین، این یافته‌ها مدل کانو را تأیید می‌کنند (Nam et al., 2021; Barzizza et al., 2024) که ویژگی‌های اساسی مانند نظافت را به‌عنوان عوامل جلوگیری‌کننده از نارضایتی طبقه‌بندی می‌کند و ویژگی‌های عملکردی مانند راحتی و منظره را به‌عنوان محرک‌های افزایش رضایت می‌شناسد. از منظر منطق غالب خدمت (Amat-Lefort et al., 2023; Tan et al., 2025)، تأکید بر تعاملات انسانی (خدمات کارکنان) نشان‌دهنده نقش هم‌آفرینی ارزش میان مهمان و هتل است که این امر با پژوهش‌های پیشین مانند صاحبی و همکاران (۲۰۲۲) هم‌راستا است و بر کشف نیازهای پنهان مشتری از طریق داده‌های بزرگ تأکید دارد.

از منظر منطق غالب خدمت^۲، برجستگی «خدمات کارکنان» به نقش محوری تعاملات انسانی در هم‌آفرینی ارزش اشاره دارد. کارکنان صرفاً ارائه‌دهنده خدمت نیستند؛ آنان تسهیل‌گر تجربه‌اند و به مهمان کمک می‌کنند تا منابع (مانند امکانات هتل) را برای دستیابی به خروجی مطلوب خود یکپارچه سازند (Tan et al., 2025). به همین قیاس، «کیفیت اتاق» و «امکانات» را می‌توان منابع عملیاتی دانست که ارزششان زمانی بالفعل می‌شود که

1. Excitement
2. Service dominant logic

در بافت تجربه شخصی مهمان به کار گرفته شوند؛ بنابراین رضایت، فرآیندی پویا و تعاملی است و هتل‌ها باید علاوه بر عرضه ویژگی‌ها، بستر هم‌آفرینی ارزش را نیز فراهم کنند (Amat-Lefort et al., 2023).

به کارگیری یادگیری ماشین توضیح‌پذیر امکان استخراج شفاف اهمیت ویژگی‌ها را فراهم کرد.

رویکرد این پژوهش کمک می‌کند علاوه بر ابعاد صریح، به کشف «ارزش مورد انتظار مشتری» و نیازهای پنهانی که معمولاً در نظرسنجی‌های سنتی مغفول می‌مانند شناسایی شده (Sahebi et al., 2022) و بدین‌سان مسیرهایی نو برای نوآوری خدمت و تمایز رقابتی گشود. استخراج نظام‌مند دیدگاه‌های جزئی از بازخورد بدون ساختار، مفهوم ارزش مورد انتظار مشتری^۱ را پالایش کرده و تصویری پویاتر از محرک‌های رضایت نسبت به روش‌های نظرسنجی ارائه می‌کند که در عمل، این نتایج زمینه مداخله‌های دقیق، تخصیص بهینه منابع و قیمت‌گذاری پویا همسو با ارزش ادراک‌شده را فراهم می‌آورد (Tan et al., 2025).

کیفیت محتوا و منظر باید هم ابعاد فیزیکی (مانند کیفیت اتاق) و هم جنبه‌های تجربی (مانند رفتار حرفه‌ای کارکنان) را بازتاب دهد. همچنین، زمینه تقسیم‌بندی دقیق و بازاریابی شخصی‌سازی شده را فراهم می‌کند؛ اقدامی که نرخ تبدیل و وفاداری را افزایش می‌دهد (Bhuiyan et al., 2023; Van Leeuwen & Koole, 2022).

از منظر طراحی تجربه، شناسایی «نیازهای پنهان» امکان بهبود هدفمند و نوآوری را می‌گشاید. هتل‌ها می‌توانند خدمات جدید بیافرینند یا خدمات موجود را چنان اصلاح کنند که ارزش ادراک‌شده مشتری به حداکثر برسد (Nam et al., 2021). برای مثال، اگر گرایش رو به رشدی نسبت به پایداری یا فناوری هوشمند دیده شود، ادغام چنین عناصری در طراحی اتاق‌ها منطقی است.

این تحقیق از مزیت الگوریتم‌های مختلف یادگیری ماشین بهره برد و عملکردشان را با معیارهایی چون دقت و اعتبار به صورت جامع سنجید تا روابط ساختاری عمیق‌تری در تجربه خدمات آشکار شود (Shrestha et al., 2024). از مشارکت‌های کلیدی، طراحی دقیق اعتبارسنجی و مهندسی ویژگی است که با ظرافت داده‌های متنی بدون ساختار

1. Expected value (EV)

سازگار می‌شود. انتخاب ویژگی پیشرفته، صرف حضور/عدم حضور را کنار گذاشته و ویژگی‌های اثرگذار را از متن‌های حجیم استخراج و اولویت‌بندی می‌کند (Wu & Riantama, 2022). پایداری مدل‌ها با معیارهایی چون میانگین مربع خطا، میانگین خطای مطلق و ضریب تعیین و اعتبارسنجی متقابل سنجیده شد تا خطر بیش‌برازش کاهش یابد و تعمیم‌پذیری افزایش پیدا کند (Alsayat, 2023).

کشف نظام‌مند «ارزش مورد انتظار مشتری» و نیازهای پنهان از طریق تحلیل متن، از جمله قوانین هم‌وقوع^۱ برای آشکارسازی دانش پنهان، فرصت‌هایی را ایجاد می‌کند که رویکردهای فرضیه‌محور غالباً از آن غافل می‌مانند (Nam et al., 2022; Sahebi et al., 2021). ترکیب متن‌کاوی و NLP در فرآیند تحلیلی، فهم اثر احساسات بر کیفیت خدمت و راهبردهای مدیریت درآمد را ژرف‌تر می‌کند و شکاف میان بازخورد خام و اقدام عملی را می‌پوشاند (Bhuiyan et al., 2023).

این پژوهش چارچوبی جامع و یکپارچه برای تحلیل داده‌محور رضایت مشتری ارائه می‌دهد که شامل پنج مرحله کلیدی است: ۱) استخراج ویژگی‌های دودویی از متن نظرات برخط؛ ۲) تحلیل همبستگی پیرسون برای روابط خطی؛ ۳) مدل‌سازی رگرسیون جنگل تصادفی برای روابط غیرخطی و اهمیت ویژگی‌ها؛ ۴) اولویت‌بندی محرک‌ها بر اساس امتیازات اهمیت و ۵) تدوین توصیه‌های عملی برای مدیریت هتل. این چارچوب، با ترکیب تحلیل‌های آماری و یادگیری ماشین، امکان درک نظام‌مند محرک‌های رضایت و تخصیص بهینه منابع را فراهم می‌سازد.

این پژوهش با محدودیت‌هایی همراه بود. نخست، تکیه بر داده‌های یک سکوی برخط، می‌تواند سوگیری انتخاب ایجاد کند؛ معمولاً تجربه‌های بسیار مثبت یا بسیار منفی بیشتر در فضای برخط بازتاب می‌یابد (Zhang & Kim, 2021) و این امر تعمیم‌پذیری یافته‌ها را به سایر سکوها یا بخش‌ها محدود می‌سازد (Tan et al., 2025). همچنین، هرچند حجم داده‌های متنی برای تحلیل کنونی کافی بود، برای بهره‌گیری بهینه از روش‌های پیشرفته یادگیری عمیق به داده‌های بزرگ‌تر نیاز است (Nam et al., 2021).

1. Association rule

از حیث روش شناختی، با وجود تحلیل اهمیت ویژگی‌ها، استخراج ویژگی تا حدی بر امتیازهای عددی از پیش تعریف شده تکیه داشت؛ رویکردی که ممکن است ظرافت‌های معنایی متن آزاد را به تمامی آشکار نکرده باشد. پیچیدگی زبان طبیعی و هم‌پوشانی واژگان میان موضوعات، تحلیل را دشوار می‌کند (Amat- ; Zhang & Kim, 2021). افزون بر این، تحلیل حاضر مقطعی است و پویایی‌های زمانی تغییر ترجیحات یا تحول کیفیت خدمت را در نظر نمی‌گیرد (Sahebi et al., 2022). گسترش دامنه داده‌ها به موقعیت‌های فرهنگی متنوع و چند سکویی در استخراج نظرات، درک دقیق‌تری از تفاوت‌های فرهنگی در انتظارات و محرک‌های رضایت فراهم می‌کند (Tan et al., 2025) و می‌تواند ترجیحات بخش‌هایی را آشکار کند که در تحلیل‌های تک‌منبعی پنهان می‌مانند (Amat-Lefort et al., 2023). فراتر از هتلداری، این روش‌شناسی در سایر بخش‌های خدماتی نیز کاربردپذیر است؛ از تحلیل ادراک «قیمت منصفانه» در سکوهای غذای برخط (Tan et al., 2025) تا تقسیم‌بندی مشتری و پیش‌بینی رفتار در بخش‌های بازاریابی گسترده‌تر (Salman et al., 2024) که همگی قابلیت تعمیم رویکردهای داده‌محور برای بهینه‌سازی خدمات را نشان می‌دهد.

تعارض منافع

هیچ گونه تعارض منافی وجود ندارد.

ORCID

Hamed Fallah Tafti



<https://orcid.org/0000-0003-4751-4394>

References

1. Adam, I. (2021). Negative tourist-to-tourist interactions, value destruction, satisfaction, and post consumption behavioral intention. *Journal of Destination Marketing & Management*, 20, Article 100557. <https://doi.org/10.1016/j.jdmm.2021.100557>
2. Alsayat, A. (2023). Customer decision-making analysis based on big social data using machine learning: A case study of hotels in Mecca. *Neural Computing and Applications*, 35(6), 4701–4722. <https://doi.org/10.1007/s00521-022-07992-x>

3. Amat-Lefort, N., Barravecchia, F., & Mastrogiacomo, L. (2023). Quality 4.0: Big data analytics to explore service quality attributes and their relation to user sentiment in Airbnb reviews. *International Journal of Quality & Reliability Management*, 40(4), 990–1008. <https://doi.org/10.1108/IJQRM-01-2022-0024>
4. Ansari, A., & Asadi, A. (2016). Assessing tourist loyalty to destinations using a data mining approach: Domestic tourists in Isfahan. *Tourism Management Studies Quarterly*, 11(35), 85–106. [In Persian]
5. Barzizza, E., Campbell, S., Ceccato, R., Dobosz, A., Haag, M., Martins, R., & Salmaso, L. (2024). A data-driven approach to understanding customer satisfaction. *Journal of Machine Intelligence and Data Science*, 5. <https://doi.org/10.11159/jmids.2024.001>
6. Benaddi, L., Ouaddi, C., Jakimi, A., & Ouchao, B. (2024). A systematic review of chatbots: Classification, development, and their impact on tourism. *IEEE Access*, 12, 78799–78810. <https://doi.org/10.1109/ACCESS.2024.3408108>
7. Bhuiyan, M., Rahman, M. A., Hoque, A. B., Ashrafuzzaman, M., & Rahman, A. (2023). Advanced analytics and machine learning for revenue optimization in the hospitality industry: A comprehensive review of frameworks. *American Journal of Scholarly Research and Innovation*, 2(2), 52–74. <https://doi.org/10.63125/8xbkma40>
8. Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G. J., Beltran, J. R., Boselie, P., Lee Cooke, F., Decker, S., & DeNisi, A. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal*, 33(3), 606–659. <https://doi.org/10.1111/1748-8583.12524>
9. Cutler, S. Q., & Carmichael, B. A. (2010). The dimensions of the tourist experience. In M. Morgan, P. Lugosi, & J. R. B. Ritchie (Eds.), *The tourism and leisure experience* (pp. 3–26). Multilingual Matters. <https://doi.org/10.21832/9781845411503-004>
10. Diwan, S. A. (2025). Optimizing guest experience in smart hospitality: Integrated fuzzy-AHP and machine learning for centralized hotel operations with IoT. *Alexandria Engineering Journal*, 116, 535–547. <https://doi.org/10.1016/j.aej.2024.11.051>
11. Elhabyb, K., Baina, A., Bellafkih, M., & Deifalla, A. F. (2024). Machine learning algorithms for predicting energy consumption in educational

- buildings. *International Journal of Energy Research*, 2024(1), Article 6812425. <https://doi.org/10.1155/2024/6812425>
12. Fallah Tafti, H. (2020). Predicting the number of tourists based on Google Trends data records using machine learning methods: A case study of tourists in Yazd. *Journal of Tourism and Development*, 10(2), 67–79. [In Persian]
 13. Fouad, A. M., Salem, I. E., & Fathy, E. A. (2024). Generative AI insights in tourism and hospitality: A comprehensive review and strategic research roadmap. *Tourism and Hospitality Research*. Advance online publication. <https://doi.org/10.1177/14673584241293125>
 14. Garner, B., Thornton, C., Luo Pawluk, A., Mora Cortez, R., Johnston, W., & Ayala, C. (2022). Utilizing text-mining to explore consumer happiness within tourism destinations. *Journal of Business Research*, 139, 1366–1377. <https://doi.org/10.1016/j.jbusres.2021.08.025>
 15. Iloranta, R., & Komppula, R. (2022). Service providers' perspective on the luxury tourist experience as a product. *Scandinavian Journal of Hospitality and Tourism*, 22(1), 39–57. <https://doi.org/10.1080/15022250.2021.1946845>
 16. Lundberg, E. (2017). The importance of tourism impacts for different local resident groups: A case study of a Swedish seaside destination. *Journal of Destination Marketing & Management*, 6(1), 46–55. <https://doi.org/10.1016/j.jdmm.2016.02.002>
 17. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
 18. Mehra, P. (2023). Unexpected surprise: Emotion analysis and aspect-based sentiment analysis (ABSA) of user generated comments to study behavioral intentions of tourists. *Tourism Management Perspectives*, 45, Article 101063. <https://doi.org/10.1016/j.tmp.2022.101063>
 19. Mirthipati, T. (2024). *Enhancing airline customer satisfaction: A machine learning and causal analysis approach* (arXiv preprint arXiv:2405.09076). <https://doi.org/10.48550/arXiv.2405.09076>
 20. Nam, S., Yoon, S., Raghavan, N., & Park, H. (2021). Identifying service opportunities based on outcome-driven innovation framework and deep learning: A case study of hotel service. *Sustainability*, 13(1), Article 391. <https://doi.org/10.3390/su13010391>

21. Prayag, G., Hosany, S., & Odeh, K. (2013). The role of tourists' emotional experiences and satisfaction in understanding behavioral intentions. *Journal of Destination Marketing & Management*, 2(2), 118–127. <https://doi.org/10.1016/j.jdmm.2013.05.001>
22. Ren, H., Pang, B., Bai, P., Zhao, G., Liu, S., Liu, Y., & Li, M. (2024). Flood susceptibility assessment with random sampling strategy in ensemble learning (RF and XGBoost). *Remote Sensing*, 16(2), Article 320. <https://doi.org/10.3390/rs16020320>
23. Sahebi, A. G., Kordheydari, R., & Aghaei, M. (2022). A new approach in marketing research: Identifying the customer expected value through machine learning and big data analysis in the tourism industry. *Asia-Pacific Journal of Management and Technology*, 2(3). <https://doi.org/10.46977/apjmt.2022v02i03.004>
24. Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, 2024, 69–79. <https://doi.org/10.58496/BJML/2024/007>
25. Sharma, R., Kumar, A., & Chuah, C. (2021). Turning the blackbox into a glassbox: An explainable machine learning approach for understanding hospitality customer. *International Journal of Information Management Data Insights*, 1(2), Article 100050. <https://doi.org/10.1016/j.ijime.2021.100050>
26. Shrestha, D., Wenan, T., Shrestha, D., Rajkarnikar, N., & Jeong, S.-R. (2024). Personalized tourist recommender system: A data-driven and machine-learning approach. *Computation*, 12(3), Article 59. <https://doi.org/10.3390/computation12030059>
27. Sun, H., Kim, M., Kim, S., & Choi, L. (2025). A methodological exploration of generative artificial intelligence (AI) for efficient qualitative analysis on hotel guests' delightful experiences. *International Journal of Hospitality Management*, 124, Article 103974. <https://doi.org/10.1016/j.ijhm.2024.103974>
28. Tan, J., Zhao, Z., Ma, W., Liu, Y., & Zhao, H. (2025). Price fairness perception on online food service platforms: A data-driven approach using fsQCA and machine learning. *International Journal of Hospitality Management*, 125, Article 104014. <https://doi.org/10.1016/j.ijhm.2024.104014>
29. Van Leeuwen, R., & Koole, G. (2022). Data-driven market segmentation in hospitality using unsupervised machine learning. *Machine Learning*

- with *Applications*, 10, Article 100414. <https://doi.org/10.1016/j.mlwa.2022.100414>
30. Veréb, V., & Azevedo, A. (2019). A quasi-experiment to map innovation perception and pinpoint innovation opportunities along the tourism experience journey. *Journal of Hospitality and Tourism Management*, 41, 208–218. <https://doi.org/10.1016/j.jhtm.2019.10.003>
31. Wang, J., Zhao, Z., Liu, Y., & Guo, Y. (2021). Research on the role of influencing factors on hotel customer satisfaction based on BP neural network and text mining. *Information*, 12(3), Article 99. <https://doi.org/10.3390/info12030099>
32. Wang, Y., Tang, T., Zhang, W., Sun, Z., & Xiong, Q. (2021). The Achilles tendon of dynamic pricing: The effect of consumers' fairness preferences on platform's dynamic pricing strategies. *Journal of Internet and Digital Economics*, 1(1), 15–35.
33. Wu, W., & Riantama, D. (2022). Determining the factors affecting customer satisfaction using an extraction-based feature selection approach. *PeerJ Computer Science*, 8, Article e850. <https://doi.org/10.7717/peerj-cs.850>
34. Wüst, K., & Bremser, K. (2025). Artificial intelligence in tourism through chatbot support in the booking process—An experimental investigation. *Tourism and Hospitality*, 6(1), Article 36. <https://doi.org/10.3390/tourhosp6010036>
35. Zhang, X., & Kim, H.-S. (2021). Customer experience and satisfaction of Disneyland hotel through big data analysis of online customer reviews. *Sustainability*, 13(22), Article 12699. <https://doi.org/10.3390/su132212699>

استناد به این مقاله: فلاح تفتی، حامد. (۱۴۰۵). تحلیل داده‌محور رضایت مشتری در صنعت هتلداری: بهره‌گیری از یادگیری ماشین برای شناسایی محرک‌های کلیدی. (e20091). فصلنامه مطالعات مدیریت گردشگری، ۲۱(۷۴)، ۲۴۸ – ۲۰۵، e20091.

doi: <https://doi.org/10.22054/tms.2025.88434.3114>



Tourism Management Studies is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.