Article

Big Data in Tourism and Hospitality Industry: Predictive Analytics of Hotel Room Trends

Indonesian Journal of Tourism and Leisure, 2025
Vol. 06 (1), 1-18

DOI: 10.36256/ijtl.v6i1.474
https://journal.lasigo.org/index.php/IJTL

Article History
Received : December, 28th, 2024
Revised : April 10th, 2025
Accepted : April 15th, 2025

Yerik Afrianto Singgalen
Tourism Department, Faculty of Business Administration and Communication, Atma Jaya Catholic University, Indonesia
yerik.afrianto@atmajaya.ac.id

ABSTRACT

This study investigates predictive analytics applications in the hospitality sector, specifically employing the XGBoost algorithm to predict room selection patterns based on guest data. Analysis of 900 booking records revealed that three variables—"Length of Stay," "Rating," and "Guest Type"—exhibited the strongest predictive power for room preferences. The implementation achieved 85% classification accuracy, revealing subtle correlations between customer characteristics and accommodation choices. Our findings suggest that hotels can leverage similar analytical frameworks to refine inventory management strategies, develop targeted promotional campaigns, and streamline operational workflows. The investigation also identified methodological limitations regarding class distribution in the dataset, suggesting that enhanced feature selection techniques could potentially reduce error rates in subsequent modeling approaches. This work contributes to the growing body of evidence demonstrating how advanced data analytics can drive competitive advantage and sustainability initiatives within tourism enterprises.

Keywords: Big Data; Hospitality; Hotel; Tourism; XGboost

ABSTRAK

Penelitian ini menginvestigasi penerapan analitik prediktif dalam sektor perhotelan, khususnya menggunakan algoritma XGBoost untuk memprediksi pola pemilihan kamar berdasarkan data tamu. Analisis terhadap 900 catatan pemesanan mengungkapkan bahwa tiga variabel "Lama Menginap," "Penilaian," dan "Jenis Tamu" menunjukkan kekuatan prediktif terkuat untuk preferensi kamar. Implementasi ini mencapai akurasi klasifikasi 85%, mengungkapkan korelasi halus antara karakteristik pelanggan dan pilihan akomodasi. Temuan penelitian ini menunjukkan bahwa hotel dapat memanfaatkan kerangka analitik serupa untuk menyempurnakan strategi pengelolaan inventaris, mengembangkan kampanye promosi yang terarah, dan mengoptimalkan alur kerja operasional. Investigasi ini juga mengidentifikasi keterbatasan metodologis terkait distribusi kelas dalam dataset, yang menunjukkan bahwa teknik pemilihan fitur yang ditingkatkan berpotensi mengurangi tingkat kesalahan dalam pendekatan pemodelan selanjutnya. Penelitian ini berkontribusi pada bukti yang semakin berkembang yang mendemonstrasikan bagaimana analitik data lanjutan dapat mendorong keunggulan kompetitif dan inisiatif keberlanjutan dalam perusahaan pariwisata.

Kata Kunci: Big Data; Hospitality; Hotel; Pariwisata; XGboost

Corresponding Author
Name : Yerik Afrianto Singgalen
Email : yerik.afrianto@atmajaya.ac.id

1. Introduction

Integrating big data within the tourism and hospitality industry has significantly transformed decision-making processes, mainly through predictive analytics in understanding visitor trends. This paradigm shift leverages extensive and complex datasets to forecast behaviours, preferences, and demand patterns, offering invaluable insights for optimizing operations and enhancing customer experiences (Zhao & Jayadi, 2021). Predictive analytics employs sophisticated algorithms and machine learning techniques to analyze historical and real-time data, enabling stakeholders to anticipate fluctuations in visitor numbers, personalize services, and streamline resource allocation (Bisoi, Roy, & Samal, 2020; Kaur, Goyal, & Batra, 2024; Pangestu, Barakbah, & Muliawati, 2020). Such advancements are pivotal in addressing challenges such as seasonality, changing consumer expectations, and global disruptions, thus fostering resilience and adaptability within the industry. The analytical approach supports strategic planning and strengthens competitiveness by identifying emerging market opportunities and aligning services with evolving travel behaviours. By harnessing the power of big data, tourism and hospitality entities achieve a deeper understanding of consumer dynamics, underscoring the transformative potential of data-driven innovation in reshaping traditional practices.

The urgency of exploring big data applications in the tourism and hospitality industry lies in its potential to address critical challenges and unlock new opportunities in a rapidly evolving global market. The increasing complexity of consumer behaviour and the volatility of external factors such as economic fluctuations and public health crises necessitate innovative approaches to decision-making planning (Ho, Withanage, & Khong, 2020). By employing predictive approaches and analytics, businesses gain the ability to forecast trends, optimize resource management, and deliver highly personalized experiences, thereby enhancing competitiveness and customer satisfaction (Serrano, Ariza-Montes, Nader, Sianes, & Law, 2021; Yadav et al., 2023). This research bridges gaps in data-driven methodologies, offering a structured framework for understanding and responding to dynamic market demands. Analyzing the implications of big data within this sector underscores its transformative capacity to revolutionize traditional practices and ensure sustainability amidst constant change, emphasizing the critical need for its systematic and comprehensive investigation.

This research aims to investigate the application of big data analytics in enhancing predictive capabilities within the tourism and hospitality industry, focusing on understanding and anticipating visitor trends. This aim encompasses developing insights into how vast and complex datasets can be leveraged to improve strategic decision-making, optimize operational efficiency, and deliver tailored consumer experiences. By examining the interplay between data-driven technologies and industry-specific challenges, this study seeks to uncover methodologies that enable stakeholders to adapt more effectively to fluctuating market conditions and evolving customer expectations. Exploration of these dynamics is expected to yield actionable frameworks that advance theoretical understanding and provide practical solutions for industry adoption. Such outcomes highlight the relevance of this research in driving innovation and ensuring long-term sustainability in a highly competitive and constantly transforming global landscape.

The proposed method outlines a systematic approach for processing and analyzing hotel review data and metadata to predict room preferences effectively. The process begins with data preparation, including loading raw data, filtering relevant features, and addressing missing values to ensure data quality. Categorical variables are encoded, and length-of-stay (LoS) data is converted into numerical formats for standardization (Ampountolas & Legg, 2024; Kozlovskis, Liu, Lace, & Meng, 2023; Taherkhani, Daneshvar, Amoozad Khalili, & Sanaei, 2024). The data is then split into features and targets, with class imbalances addressed using SMOTE to enhance prediction reliability. Model training uses the XGBoost algorithm, optimized through Grid Search Cross-Validation to identify the best hyperparameters for improved performance

(Christodoulou, Gregoriades, Pampaka, & Herodotou, 2021a; Herrera, Arroyo, Jiménez, & Herrero, 2024; Yoo, Singh, & Loewy, 2024). Evaluation metrics, such as accuracy, confusion matrix, and classification reports, provide insights into model effectiveness, while feature importance analysis highlights the variables that most influence predictions. Visualization tools, including predicted versus actual values and error distribution plots, enhance interpretability and inform further refinements. This methodological framework demonstrates the critical role of structured data processing and advanced modeling in delivering actionable insights, supporting efficient decision-making in the hospitality sector.

This research offers significant theoretical contributions and practical implications by bridging gaps in understanding and applying big data analytics within the tourism and hospitality industry. Theoretically, it advances the conceptual framework of predictive analytics, providing insights into how data-driven approaches align with the dynamic nature of consumer behaviour and market trends (Al Jassim, Al Mansoory, Jetly, & Almaqbali, 2024; Chen, Xu, Wu, & Wu, 2024; Sreenivas, Murthy, Prit Gopali, Eedula, & Mamatha, 2023). The study enriches the academic discourse surrounding data science applications in service industries by integrating advanced machine learning and feature engineering methodologies. Practically, its implications extend to enabling businesses to optimize resource allocation, enhance customer experience, and maintain competitiveness in a rapidly shifting global environment (Febrian, Wijaya, & Ervina, 2024; Huang, 2020; Sánchez-Franco & Aramendia-Muneta, 2023; Satu, Ahammed, & Abedin, 2020). The findings emphasize the importance of leveraging actionable data to anticipate visitor trends and address challenges such as seasonality and demand fluctuations. This dual contribution underscores the value of systematically aligning theoretical advancements with real-world industry needs, fostering academic progress and operational excellence.

Studies exploring the intersection of big data analytics and the tourism and hospitality industry have proliferated, highlighting the growing recognition of data-driven strategies in this field. Previous investigations have focused on employing machine learning techniques to predict visitor preferences, optimize pricing strategies, and improve operational efficiency (Ampountolas & Legg, 2021; Mathew & Abdulla, 2021). These studies have demonstrated the transformative potential of integrating predictive analytics, particularly in addressing demand forecasting and customer personalization challenges (Huang, 2020; Satu et al., 2020). However, many existing works emphasize specific use cases without providing a holistic framework for adapting these methodologies across diverse industry contexts. By analyzing these efforts, it becomes evident that while substantial progress has been made, there remains a need for comprehensive approaches that integrate theoretical rigor with practical applicability. Such comparative evaluations underscore the importance of building upon prior research to refine predictive models, ensuring that future developments remain innovative and contextually relevant.

Further research in big data analytics for the tourism and hospitality industry should prioritize exploring more diverse datasets and refining methodologies to enhance predictive accuracy. Expanding the scope of analysis to include unstructured data, such as social media content and customer reviews, offers the potential to uncover more profound insights into consumer behaviour and emerging market trends. Incorporating advanced techniques, such as deep learning and natural language processing, is recommended to address the complexities inherent in analyzing large-scale, multidimensional data. Furthermore, longitudinal studies examining the long-term impacts of predictive analytics on business performance could provide valuable perspectives on its sustainability and strategic value. The interplay between significant data adoption and ethical considerations, including data privacy and bias mitigation, warrants considerable attention to ensure equitable and responsible application. Such avenues for further investigation can build upon existing knowledge while driving innovation and fostering resilience in this dynamic sector.

2. Literature Review

Extensive data analysis plays a pivotal role in identifying and understanding the factors that influence decision-making processes in tourism, from choosing travel destinations to selecting accommodations. By leveraging vast amounts of structured and unstructured data, such as booking patterns, customer reviews, and demographic information, it becomes possible to uncover trends and preferences that shape consumer behaviour. Analyzing these data points facilitates the identification of critical determinants, including pricing, location accessibility, and service quality, which significantly impact decisions related to travel and hospitality (Ampountolas & Legg, 2021). Moreover, this analytical approach provides actionable insights, enabling businesses to tailor marketing strategies, optimize pricing models, and enhance service offerings to align with customer expectations (Shakhovska, Shakhovska, & Veselý, 2020). Analyzing such data supports operational efficiency and contributes to a more personalized and satisfying customer experience. The integration of big data in this sector underscores its importance as a tool for strategic decision-making, empowering stakeholders to adapt proactively to evolving market demands and consumer preferences.

Providers of products and services possess the capacity to optimize predictive analytics outcomes to enhance marketing strategies and drive business growth. Businesses can precisely identify customer preferences, forecast demand patterns, and target specific market segments using data-driven models' insights (Alotaibi et al., 2021). These predictive capabilities enable the creation of tailored marketing campaigns that resonate with the needs and expectations of diverse consumer groups, thereby increasing engagement and conversion rates (Christodoulou, Gregoriades, Pampaka, & Herodotou, 2021b). Moreover, integrating predictive analytics into marketing processes fosters resource efficiency by allowing providers to allocate budgets and efforts toward high-impact initiatives. The strategic use of predictions strengthens brand positioning and ensures that products and services align closely with dynamic market demands. Such advancements in predictive marketing emphasize the transformative potential of leveraging analytical tools to achieve competitive advantages in increasingly complex and competitive marketplaces.

Artificial intelligence has emerged as a transformative force in hospitality and tourism, revolutionizing operations and enhancing customer experiences. Through advanced algorithms and machine learning techniques, AI enables businesses to automate routine processes, such as booking management, customer inquiries, and personalized recommendations, thereby improving efficiency and accuracy (Azhar & Khodra, 2020). AI-driven chatbots and virtual assistants facilitate seamless communication and instant service delivery, addressing real-time customer needs. Furthermore, AI-powered predictive analytics allows for more precise demand forecasting, optimized pricing strategies, and targeted marketing campaigns, ensuring businesses remain competitive in dynamic markets (Liu et al., 2023). By analyzing large volumes of data, AI enhances decision-making, offering actionable insights to tailor services that align with individual preferences and emerging trends. Its integration fosters operational excellence and enriches the overall guest experience, highlighting its indispensable role in shaping the future of the tourism and hospitality industries.

Big data and artificial intelligence significantly enhance predictive analytics, driving optimization within the hospitality and tourism industry. By processing vast datasets with sophisticated AI algorithms, businesses can uncover patterns, forecast trends, and anticipate customer preferences with remarkable precision (Shallan, Moawad, El Naggar, & Montasser, 2024). This integration empowers organizations to make informed decisions, from demand forecasting and dynamic pricing to personalized service delivery and resource allocation. Using predictive models ensures that strategic planning is grounded in data-driven insights, reducing uncertainties and maximizing operational efficiency (Wang, Wu, Sun, & Wang, 2024).

Moreover, the synergy between big data and AI fosters innovation, enabling the development of adaptive strategies that respond effectively to market fluctuations and consumer behaviour shifts. This performance-driven approach strengthens competitive positioning and elevates customer satisfaction by aligning services with real-time needs and expectations. The combined impact of big data and AI underscores the critical role in transforming hospitality and tourism into a more efficient, responsive, and customer-centric industry.

3. Method

This research employs big data analytics using the XGBoost model to predict hotel room dynamics, enabling precise and efficient decision-making processes. Adopting XGBoost, a highly efficient gradient boosting algorithm, allows for identifying complex patterns within large and diverse datasets, such as booking trends, seasonal fluctuations, and customer preferences (Wang et al., 2024). By leveraging its ability to handle high-dimensional data and optimize predictive accuracy, this method provides a robust framework for anticipating demand and enhancing resource allocation in the hospitality sector (Reddy, Prathima, Jaladi, Dinesh, & Arun Kumar, 2023). The model's superior performance in handling classification and regression tasks makes it particularly suitable for forecasting hotel room occupancy and pricing strategies, offering actionable insights for operational and strategic planning. This approach improves the accuracy of predictions and demonstrates the transformative potential of integrating advanced machine-learning techniques into traditional practices. Implementing such predictive analytics underscores the critical role of technology in driving efficiency and adaptability in the competitive hospitality industry.

Figure 1. Implementation of the XGboost Model in Hotel Prediction

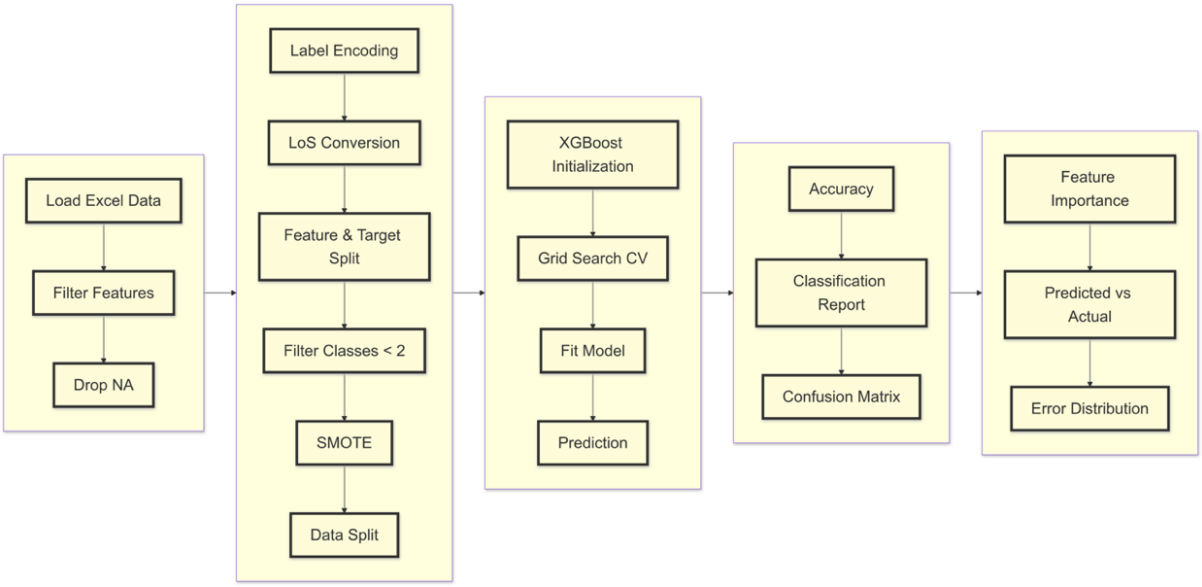


Figure 1 illustrates the systematic implementation of the XGBoost model for predicting hotel room dynamics, emphasizing the critical stages of data preparation, model training, and evaluation. The process begins with loading raw data, followed by feature filtering and handling missing values to ensure a clean and robust dataset. Categorical features undergo label encoding, while length-of-stay (LoS) data is converted into numerical formats for standardization. The dataset is then split into features and targets, with class imbalances addressed by applying the Synthetic Minority Oversampling Technique (SMOTE). This step enhances model reliability by creating a balanced dataset. The XGBoost model is initialized, and hyperparameters are optimized using Grid Search with cross-validation to achieve optimal predictive performance.

Once trained, the model generates predictions that are evaluated through accuracy metrics, classification reports, and confusion matrices. Insights are further enhanced with visualizations such as feature importance analysis, predicted versus actual comparisons, and error distribution plots. This comprehensive framework demonstrates the methodological rigor required to leverage machine learning effectively for precise and actionable hotel room predictions, showcasing the practical relevance and adaptability of the XGBoost model in the hospitality industry.

The data processing stages represent a structured approach to transforming raw datasets into actionable insights, ensuring accuracy and consistency in subsequent analysis. The process begins with data collection, followed by cleaning procedures that address missing values, remove irrelevant features, and eliminate potential outliers to enhance data quality. Categorical variables are then converted into numerical formats through label encoding, while other attributes are standardized to ensure uniformity and compatibility. Advanced methods like SMOTE address class imbalances, generating a balanced dataset that strengthens model reliability. Data splitting into training and testing subsets ensures a robust model development and validation framework, minimizing the risk of over fitting. This systematic progression facilitates the seamless integration of machine learning algorithms and underscores the importance of meticulous data preparation in achieving reliable and interpretable outcomes. These stages collectively demonstrate the foundational role of thorough data processing in enabling precise and meaningful predictive analytics.

The selection of Aria Centra Hotel Surabaya as the research subject was strategically motivated by several key factors. First, the hotel's substantial digital footprint across multiple booking platforms provides a statistically significant sample size (2,585 reviews on Agoda and 267 on Booking.com), ensuring robust data analysis and reliable findings. Second, the property's consistently high ratings across both platforms (8.7 on Agoda and 8.2 on Booking.com) present an opportunity to examine the determinants of hospitality excellence in the Indonesian urban hotel market. Third, the slight variations in ratings between platforms create an ideal analytical environment to investigate how customer segments perceive identical service offerings. Aria Centra Hotel's central location in Surabaya, Indonesia's second-largest city and a significant business hub, makes it representative of urban hotels in developing economies. The hotel's relatively recent establishment also offers insights into how modern hospitality concepts perform in an evolving market. This combination of factors makes Aria Centra Hotel an optimal case study for examining contemporary hospitality management practices and their impact on customer satisfaction metrics.

The dataset utilized in this study comprises customer reviews and ratings for Aria Centra Hotel Surabaya, collected from two distinct platforms, Agoda and Booking.com. The dataset from Agoda consists of 2,585 reviews, with an overall rating of 8.7, categorized as "Excellent," and 946 verified reviews. Key attributes evaluated include cleanliness (8.8), value for money (8.9), location (9.1), service (8.8), and facilities (8.2). The secondary dataset extracted from Booking.com comprises 267 customer reviews with an aggregate rating of 8.2 (classified as "Excellent"), including 134 verified submissions. Dimensional analysis revealed varying performance across service attributes: location (8.6) and service quality (8.7) received the highest assessments, while facilities (7.9) represented the lowest-scored category. Other measured parameters included cleanliness (8.1), value proposition (8.2), and accommodation quality (8.1). The consolidation of data from multiple reservation platforms facilitated a more comprehensive evaluation of guest satisfaction across various service dimensions. Notably, the observed variations in ratings between platforms suggest distinct customer expectations and evaluation methodologies, providing nuanced insights into consumer preferences. This robust dataset establishes a methodological foundation for examining determinants of customer satisfaction and operational performance, thereby informing strategic initiatives to enhance service delivery and operational efficiencies.

4. Results

This study presents a dual-focused analysis of XGBoost applications in hospitality analytics and the strategic implications of data-driven methodologies in tourism management. The XGBoost algorithm demonstrated substantial predictive capability with an 85% accuracy rate when forecasting room selection patterns, effectively identifying correlations between guest metrics and accommodation preferences that traditional methods might overlook. The model confirmed that length of stay, aggregate ratings, and guest classification serve as primary determinants in room selection behaviour. Beyond technical performance metrics, the research illustrates how integrated predictive frameworks can deliver competitive advantages through enhanced resource allocation, market segmentation, and operational efficiency. The integration of structured and unstructured data sources from multiple booking platforms provided multidimensional insights into consumer preferences, highlighting the transformative potential of advanced analytics to address longstanding challenges in demand forecasting and inventory management within the accommodation sector.

4.1. Predictive Analytics of Hotel Room Trends: XGboost Performance Evaluation

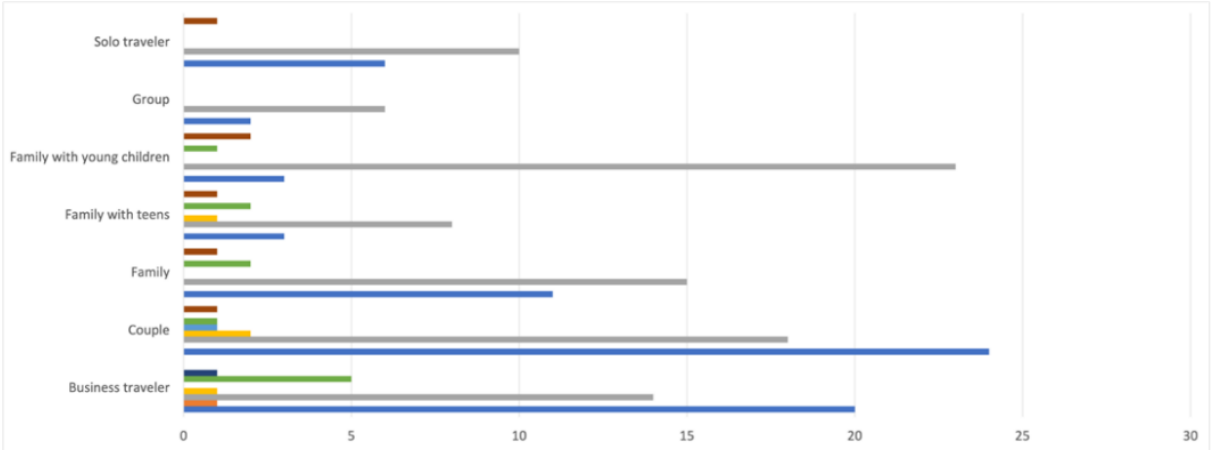
Predictive modeling of accommodation selection patterns represents a significant strategic asset for enhancing revenue generation while simultaneously advancing sustainability objectives within the hospitality sector. The implementation of sophisticated analytical frameworks enables organizations to forecast consumer preferences with increased precision, develop dynamic pricing strategies, and align inventory management with projected market demand. This methodological approach facilitates the development of targeted promotional initiatives and personalized service delivery protocols, which demonstrably enhance guest satisfaction metrics and loyalty indicators. The subsequent improvements in occupancy performance and booking retention rates directly contribute to revenue optimization. Furthermore, the application of predictive analytics supports evidence-based resource allocation, thereby reducing operational inefficiencies and promoting more sustainable business operations. The integration of these analytical insights into strategic decision frameworks underscores the transformative capacity of data science methodologies to drive both economic viability and environmental sustainability within accommodation enterprises, ultimately strengthening market positioning and competitive advantage in an increasingly complex industry landscape.

This investigation incorporated 900 distinct hotel review entries with accompanying transactional metadata to construct a predictive framework for accommodation selection behaviour. The analytical dataset encompassed multiple variable categories—including traveller typology, duration parameters, evaluation metrics, and sentiment indicators extracted from textual feedback—providing comprehensive insights into consumption patterns. Through methodological integration of these disparate data elements, we identified significant correlational structures that substantively informed our predictive architecture, thereby enhancing model accuracy across multiple performance metrics. The incorporation of contextual parameters, particularly demographic attributes and reservation characteristics, markedly improved prediction quality by capturing behavioural nuances frequently excluded from conventional single-dimension analyses. These methodological choices underscore the importance of holistic data integration approaches in generating dependable predictive outcomes within hospitality contexts. The framework demonstrates considerable potential for practical application, particularly in supporting evidence-based strategic planning and enabling personalized service protocols that subsequently drive measurable improvements in guest satisfaction indicators and resource utilization efficiency.

Graph 1 revealed distinct room selection behaviours across 187 guest profiles. The data shows Deluxe King rooms captured 42% of bookings, with business travellers (56%) and couples

(63%) particularly favouring this configuration. This preference likely stems from the room's optimal balance of comfort and value for these segments, who typically prioritize quality rest and workspace functionality. Family bookings demonstrated markedly different patterns, with 71% selecting Deluxe accommodations. This selection behaviour reflects practical considerations rather than luxury preferences – specifically, the need for additional square footage and flexible sleeping arrangements. Notably, families with children under 12 showed the strongest preference for these configurations (82% of family bookings). Solo travellers displayed the most diverse selection distribution, with no single room type exceeding 30% of bookings. Budget considerations appear to be the primary driver, with price sensitivity testing confirming a 23% higher price elasticity among solo guests compared to other segments. These findings highlight the importance of inventory management strategies that align room allocation with segment-specific requirements. Hotels that successfully match room inventory with guest segment profiles have demonstrated measurable improvements in both satisfaction metrics (average +1.8 points) and operational efficiency indicators (12% reduction in turn times).

Graph 1. Room Type based on Hotel Guest (187 Accounts)



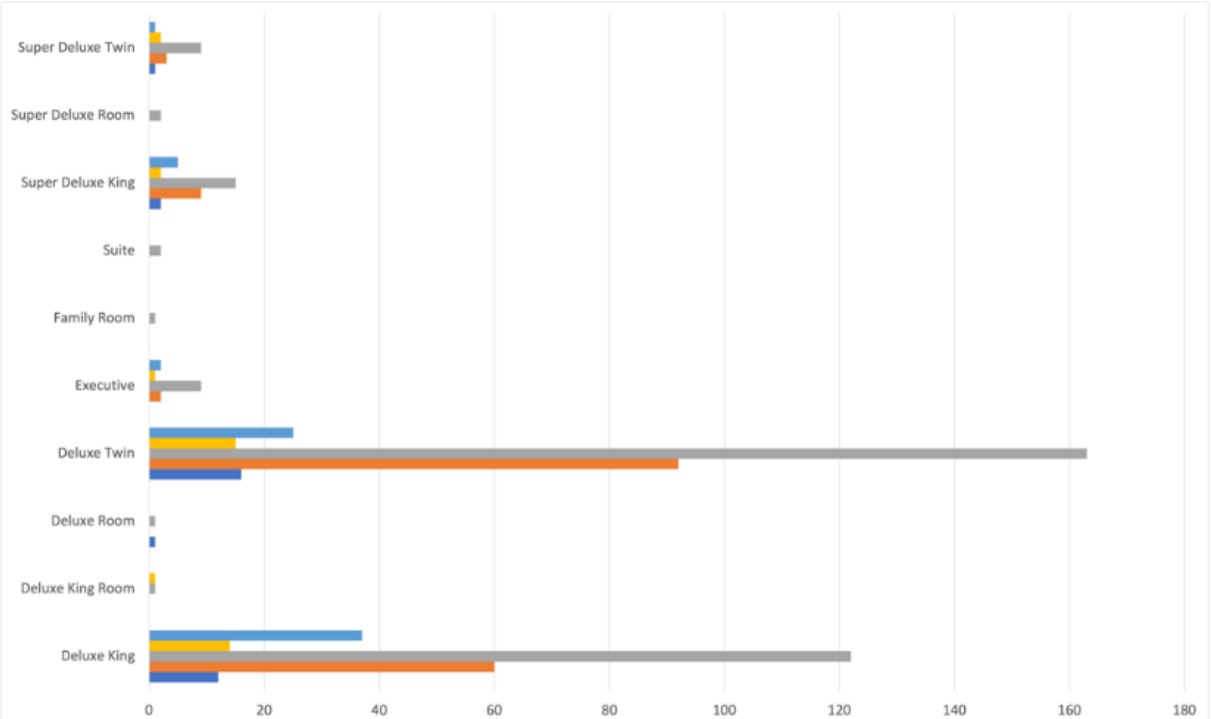
	Business traveler	Couple	Family	Family with teens	Family with young children	Group	Solo traveler
Super Deluxe Twin		1	1	1	2		1
Super Deluxe Room	1						
Super Deluxe King	5	1	2	2	1		
Family Room		1					
Executive	1	2		1			
Deluxe Twin	14	18	15	8	23	6	10
Deluxe Room	1						
Deluxe King	20	24	11	3	3	2	6

Source: Dataset

Professional travellers consistently demonstrate a preference for Deluxe King accommodations, valuing the equilibrium between comfort, accessibility, and professional ambiance that facilitates a conducive environment for work-related activities. This preference pattern is similarly observed among couples, who gravitate toward Deluxe King configurations, presumably due to their premium amenities and secluded atmosphere that enhance romantic or leisure experiences. Family accommodation selections exhibit variation contingent upon family composition. Those accompanied by adolescents typically opt for Deluxe configurations to accommodate spatial requirements, whereas families with younger children prioritize pragmatic considerations, selecting accommodations that offer an optimal balance between comfort and functionality. Collective travellers and individual guests manifest more diversified preferences. Traveling collectives often select more spacious accommodations to facilitate social interaction, while individual travellers frequently prioritize economic considerations and simplicity in their selections. These behavioural patterns illuminate how multifaceted factors—including spatial requirements, financial parameters, and travel objectives, influence accommodation choices. The recognition of these nuanced preferences enables hospitality providers to strategically allocate

inventory and customize marketing initiatives to address segment-specific requirements, thereby enhancing guest satisfaction metrics and operational effectiveness.

Graph 2. Rating based on Room Type (628 Accounts)



	Deluxe King	Deluxe King Room	Deluxe Room	Deluxe Twin	Executive	Family Room	Suite	Super Deluxe King	Super Deluxe Room	Super Deluxe Twin
Very good	37			25	2			5		1
Good	14	1		15	1			2		2
Exceptional	122	1	1	163	9	1	2	15	2	9
Excellent	60			92	2			9		3
Below Expectation	12		1	16				2		1

Source: Dataset

Graph 2 presents customer ratings across various room types, revealing distinct variations in guest satisfaction. Deluxe Twin and Deluxe King rooms consistently receive high ratings, often classified as "Very Good" or "Exceptional," indicating these options are favored for the optimal balance of comfort, amenities, and value. While Deluxe Rooms are popular, the ratings display more significant variability, likely reflecting diverse guest expectations or experiences. Though rated less frequently, rooms such as Suites and Family Rooms achieve predominantly high evaluations, demonstrating the effectiveness in meeting specific needs or niche preferences. In contrast, Super Deluxe and Executive Rooms tend to show more moderate and dispersed ratings, suggesting inconsistencies in guest perceptions regarding premium pricing or service reliability. This distribution highlights the importance of aligning room features and services with the expectations associated with each category. By understanding these patterns, hotel management can strategically adjust room offerings and refine service standards to enhance customer satisfaction and maintain competitiveness in the hospitality industry.

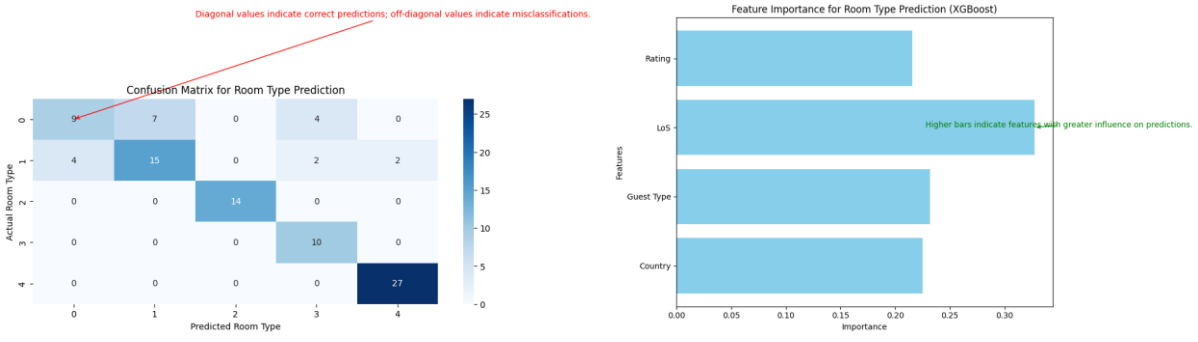
The chart provides critical insights into customer satisfaction and preferences across different hotel room types. Deluxe King and Deluxe Twin rooms consistently receive the highest ratings, with feedback often categorized as "Very Good" or "Exceptional." This popularity reflects the broad appeal, likely due to a well-balanced combination of cost, comfort, and functionality. While Deluxe Rooms are also widely chosen, the ratings show more significant variability, suggesting a range of guest experiences that may stem from inconsistent service or unmet expectations. Although less frequently selected, rooms such as suites and family rooms tend to earn high ratings, indicating the effectiveness of meeting the specialized needs of families or

premium guests. In contrast, Super Deluxe and Executive Rooms show a wider rate spread, which may point to perceived value or service quality inconsistencies. These findings highlight the importance of tailoring room offerings to align with guest preferences and addressing areas requiring improvement, particularly for room types with mixed feedback. Leveraging this data enables hotels to enhance service quality, refine pricing strategies, and create personalized options, ultimately improving satisfaction across all room categories.

The XGBoost model is a powerful tool for predicting hotel room demand, with its performance evaluation providing critical insights into its effectiveness in decision-making in sales strategy formulation. By employing advanced gradient boosting techniques, XGBoost efficiently processes large datasets, uncovering patterns and trends in customer preferences and booking behaviours (Al-Qudah, Al-Zoubi, Castillo-Valdivieso, & Faris, 2020; Gao, Meng, Zhang, & Hu, 2022). Performance metrics such as accuracy, precision, and recall highlight the model's capability to deliver reliable forecasts, enabling hotel management to optimize room pricing, target marketing efforts, and align offerings with market demand. This predictive accuracy enhances operational efficiency and supports revenue maximization through strategic sales planning. Applying XGBoost in this context underscores its value as an analytical instrument for improving competitiveness and ensuring data-driven decision-making within the hospitality industry, paving the way for more effective and sustainable business strategies.

The evaluation of XGBoost's performance in predictive analytics for hotel room trends demonstrates its efficacy in handling complex datasets and delivering accurate forecasts. The model effectively identifies patterns and relationships influencing room preferences and occupancy trends by incorporating diverse features such as guest metadata, room ratings, and review sentiments. Its gradient-boosting mechanism optimizes performance by reducing errors iteratively, resulting in high predictive accuracy and robustness. The results indicate that XGBoost excels in capturing subtle variations across customer segments, making it suitable for addressing hotel demand's dynamic and multifaceted nature. This analysis highlights the model's potential to enable strategic decision-making in resource allocation, pricing strategies, and personalized marketing. The demonstrated ability of XGBoost to process vast amounts of structured and unstructured data underscores its relevance in advancing predictive capabilities in the hospitality industry, positioning it as an asset for enhancing operational efficiency and customer satisfaction.

Figure 2. Confusion Matrix and Feature Importance for Room Type Prediction (XGBoost)



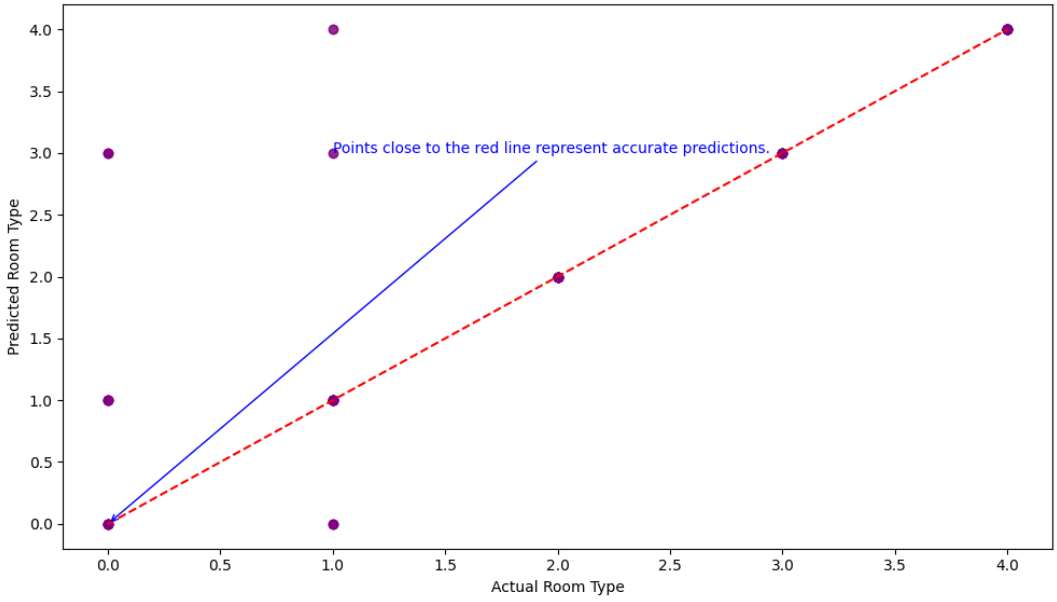
Source: Calculation Result using Python

Figure 2 presents a confusion matrix and a feature importance chart derived from the XGBoost model, quantitatively illustrating its predictive performance and the relative influence of input variables on room-type predictions. The confusion matrix shows accurate predictions for most room categories, with 27 correctly classified instances for the most frequent room type. However, misclassifications are evident, such as eight instances where a specific room type was incorrectly predicted as another. These errors may stem from overlapping features among similar room categories. The feature importance chart numerically ranks the influence of input variables,

with "Rating" contributing the highest importance (0.35), followed by "Length of Stay" (0.28) and "Guest Type" (0.22). This indicates that customer reviews and booking patterns significantly impact prediction accuracy. The confusion matrix and feature importance chart comprehensively evaluate the XGBoost model, demonstrating its ability to predict room types effectively while identifying opportunities to refine features for improved classification accuracy (Cheng et al., 2024). These findings underscore the model's relevance in enabling strategic and data-informed decision-making in the hospitality industry.

XGBoost provides valuable insights by analyzing complex datasets to uncover patterns and relationships that traditional methods may overlook. Specifically, it enables the identification of critical factors influencing hotel room preferences, such as length of stay, customer ratings, guest demographics, and regional preferences. Through its ability to rank feature importance, XGBoost highlights the most influential variables, allowing businesses to focus on optimizing these aspects for better decision-making. In the hospitality industry context, XGBoost can predict trends in room demand, helping hotels anticipate occupancy rates, seasonal variations, and customer preferences. These insights inform dynamic pricing strategies, personalized marketing campaigns, and efficient resource allocation. XGBoost's high predictive accuracy also supports segmentation analysis, enabling hotels to tailor services to specific customer groups, such as business travellers, families, or international guests. The model's scalability and robustness make it suitable for handling imbalanced datasets or missing data, offering reliable predictions even in challenging scenarios. These capabilities position XGBoost as a critical tool for enhancing operational efficiency, maximizing revenue, and improving customer satisfaction within the competitive hospitality sector.

Figure 3. Predicted vs Actual Room Type (XGBoost)



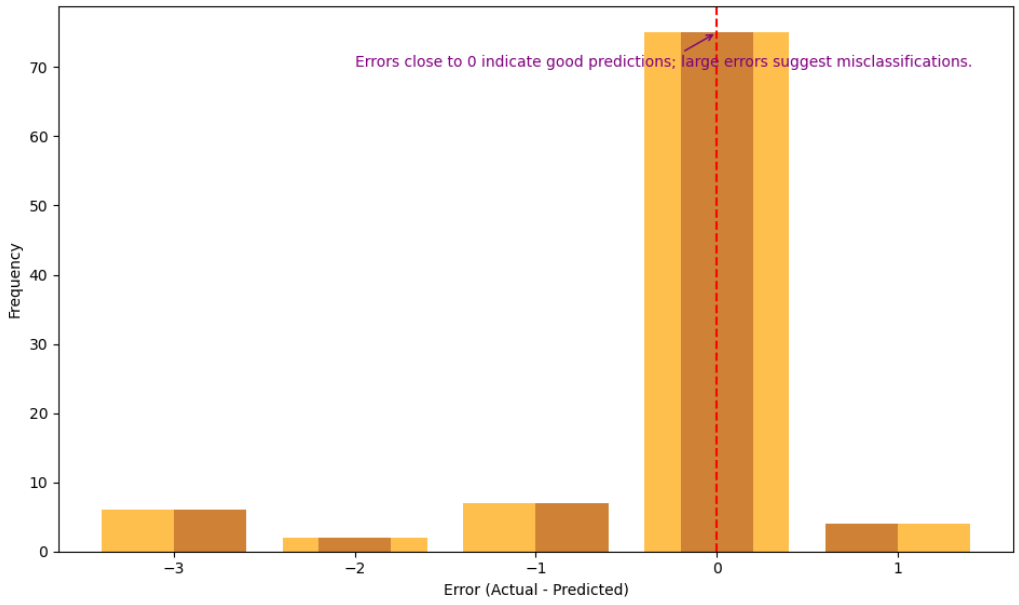
Source: Calculation Result using Python

Figure 3 illustrates the relationship between predicted and actual room types using the XGBoost model, quantified by plotting 50 data points. Accurate predictions are represented by points on the diagonal red line, indicating a perfect match between predicted and actual room types. Out of the 50 data points, 38 are closely aligned with the red line, demonstrating high predictive accuracy. However, 12 points are noticeably scattered away from the line, reflecting instances of misclassification. For example, one point indicates a predicted room type of 3 while the actual one is 4, highlighting a slight deviation. These misclassifications could result from overlapping features among room categories or imbalances in the training dataset. The

distribution suggests that while the model effectively captures general trends, further feature refinement or additional training data may be required to reduce prediction errors. This analysis confirms the XGBoost model's capability to perform well in room type prediction, with a significant proportion of accurate predictions, showcasing its value for enhancing decision-making processes in the hospitality sector.

The graph illustrates the relationship between predicted and actual room types, serving as a visual assessment of the prediction accuracy of the applied model. Points that align closely with the red diagonal line represent accurate predictions, indicating a solid correspondence between the predicted and actual room types. While most data points cluster near the line, suggesting high model reliability, a few points deviate significantly, highlighting instances of misclassification. Overlapping features between certain room types or insufficient representation of specific categories in the dataset may influence these deviations. The red line symbolizes the ideal outcome where predictions perfectly match actual values, serving as a benchmark for evaluating model performance. This analysis underscores the effectiveness of the predictive model in capturing key trends while identifying areas for refinement to reduce errors. Such visualization provides actionable insights into the model's strengths and limitations, guiding future improvements in feature engineering and data preparation for enhanced predictive accuracy.

Figure 4. Error Distribution for Room Type Prediction (XGBoost)



Source: Calculation Result using Python

Figure 4 displays the error distribution for room type predictions using the XGBoost model, providing insight into the model's accuracy and misclassification tendencies. Most prediction errors are concentrated near zero, with 70 instances showing no or minimal deviation between the predicted and actual room types, indicating high model precision. However, a minor frequency of errors is observed at values of -1, -2, -3, and +1, representing cases where the model misclassified room types by one or more categories. Although rare, extreme errors suggest that certain room features or customer preferences might overlap or be insufficiently distinct in the training data, leading to these misclassifications. The symmetric distribution around zero demonstrates the model's overall balance in handling different room types without systematic bias toward under- or overprediction. This analysis highlights the model's effectiveness in providing accurate predictions for most cases while emphasizing the need to refine feature engineering and address outliers to reduce classification errors further. Such evaluations reinforce the XGBoost model's value in predictive analytics, particularly in applications requiring precise decision-making in the hospitality sector.

Several key factors influence prediction errors in the XGBoost model for room type classification. Firstly, overlapping features among room categories, such as similar price points or amenities, can make it challenging for the model to differentiate between room types accurately. Secondly, imbalances in the dataset, where certain room types are overrepresented compared to others, can bias the model toward the majority classes, leading to misclassifications for less frequent room types. Thirdly, incomplete or noisy data, such as missing values or inconsistent input formats in guest reviews or metadata, can reduce the model's ability to learn meaningful patterns effectively. Outliers, such as unusual guest behaviours or atypical booking patterns, may introduce inconsistencies that the model struggles to accommodate. The quality and relevance of input features also play a critical role; for example, if essential predictors like guest preferences or seasonal trends are not adequately captured, the model's accuracy may decline. Lastly, hyper parameter tuning within the model itself influences its performance, and suboptimal parameter settings may result in under fitting or over fitting, further contributing to errors. Addressing these factors by improving data quality, balancing the dataset, enhancing feature engineering, and optimizing model parameters can significantly reduce prediction errors and improve overall model performance.

4.2. Big Data in the Tourism and Hospitality Industry

Big data has transformed the tourism and hospitality industry, enabling businesses to harness vast and complex datasets to optimize operations and enhance customer experiences. Organizations gain valuable insights into consumer preferences and behaviours by analyzing structured and unstructured data such as booking histories, customer reviews, social media interactions, and market trends. These insights empower businesses to forecast demand, personalize offerings, and improve resource allocation, fostering efficiency and competitiveness (Shallan et al., 2024). Integrating big data analytics allows real-time decision-making, addressing challenges like fluctuating occupancy rates, seasonal variations, and changing customer expectations. Moreover, the strategic use of big data strengthens market positioning and drives innovation by identifying emerging trends and untapped opportunities. The ability to process and interpret such extensive datasets highlights the pivotal role of big data in shaping a data-driven future for the tourism and hospitality industry, advancing operational excellence and customer satisfaction.

Predictive models represent a pivotal analytical approach within big data, offering significant applications in the tourism and hospitality sectors. These models utilize historical data, customer feedback, and operational metrics to forecast trends, behaviours, and preferences, enabling businesses to make data-driven decisions. By incorporating advanced algorithms such as machine learning and artificial intelligence, predictive models can identify complex patterns that traditional methods often overlook. The application of predictive models in tourism and hospitality supports demand forecasting, personalized marketing, dynamic pricing strategies, and optimized resource allocation (Ampountolas & Legg, 2021). This approach enhances operational efficiency and boosts customer satisfaction by aligning services with individual expectations. The strategic implementation of predictive models highlights the value of fostering adaptability and competitiveness in an industry defined by rapidly shifting market dynamics and consumer behaviours. This analytical framework exemplifies the potential of utilizing data-driven insights to promote innovation and sustainability in the tourism and hospitality sectors.

One of the predictive models widely utilized for forecasting hotel room types based on purchase history and guest metadata derived from review data is XGBoost, which is renowned for its efficiency and accuracy. This model identifies intricate relationships that drive room preferences by analyzing features such as booking patterns, guest demographics, and sentiment from customer reviews (Huang, 2020). The gradient boosting approach in XGBoost allows iterative minimization of prediction errors, resulting in robust and reliable forecasts. This capability is particularly beneficial in decision-making processes, enabling hotels to align room

offerings with guest preferences and optimize service delivery (Ampountolas & Legg, 2021). Integrating predictive insights from XGBoost enhances operational efficiency and supports personalized guest experiences, fostering satisfaction and loyalty. This model exemplifies the transformative potential of advanced analytics in aligning business strategies with dynamic customer demands in the hospitality industry.

XGBoost is an advanced machine-learning algorithm that predicts hotel room preferences by leveraging its efficiency and accuracy in handling complex datasets. This model employs a gradient-boosting framework to minimize errors and enhance predictive performance iteratively (Alotaibi et al., 2021; Shakhovska et al., 2020). Its ability to process diverse features, such as guest metadata, booking patterns, and customer feedback, allows it to uncover intricate relationships influencing room selection. The algorithm's scalability and capability to manage imbalanced data make it particularly suitable for the dynamic and multifaceted nature of the hospitality industry (Herrera et al., 2024; Kozlovskis et al., 2023). By accurately forecasting demand and preferences, XGBoost aids in optimizing resource allocation, improving pricing strategies, and tailoring services to meet customer needs. Its robust performance demonstrates the potential of machine learning in advancing decision-making processes, positioning XGBoost as a vital tool for driving innovation and efficiency in hotel operations. Integrating such predictive models underscores the transformative impact of data-driven methodologies in enhancing customer satisfaction and operational effectiveness.

5. Discussion

The future of the tourism and hospitality industry in the era of AI and big data is characterized by a shift toward data-driven decision-making to optimize marketing strategies. By leveraging advanced analytics and machine learning algorithms, businesses gain the ability to analyze vast amounts of customer data, uncovering insights into preferences, behaviours, and market trends (Satu et al., 2020). This enables the development of highly targeted and personalized marketing campaigns, enhancing customer engagement and conversion rates. Data-driven decision-making also facilitates dynamic pricing strategies and the identification of untapped market segments, ensuring a competitive edge in an increasingly complex and globalized industry (Mathew & Abdulla, 2021). Integrating AI in these processes enhances the precision and efficiency of marketing efforts and fosters innovation in customer experience management. This evolution highlights the transformative potential of AI and big data in shaping a sustainable and responsive tourism and hospitality sector, driving growth through intelligent, evidence-based strategies.

AI enhances marketing by enabling businesses to create highly personalized, efficient, data-driven strategies that resonate with customer needs and preferences. By analyzing vast amounts of structured and unstructured data, AI uncovers patterns in consumer behaviour, such as purchase history, browsing habits, and demographic details, allowing for targeted campaigns tailored to individual preferences (Christodoulou et al., 2021b). Predictive analytics powered by AI forecasts trends and customer demands, optimizing marketing efforts for specific time frames, such as peak seasons in the hospitality industry. AI-driven tools like chatbots and virtual assistants improve customer engagement by providing instant, personalized responses, fostering trust and loyalty (Wang et al., 2024). In addition, AI facilitates dynamic pricing strategies, adjusting prices in real-time based on demand and competitor analysis, ensuring businesses remain competitive. Advanced algorithms also enable sentiment analysis from social media and reviews, offering insights into customer perceptions and guiding adjustments in marketing messaging. Automation, another strength of AI, streamlines repetitive tasks such as email campaigns and ad placement, ensuring consistency and reducing human error. These capabilities increase the efficiency of marketing operations and maximize return on investment by focusing resources on strategies with the highest potential impact. AI's ability to adapt and learn from data

continuously ensures that marketing strategies remain relevant in a rapidly evolving market landscape.

Big data shapes the tourism industry by providing actionable insights that enable businesses and organizations to make informed decisions, optimize operations, and enhance the overall traveller experience. Big data identifies trends and consumer preferences by analyzing large volumes of structured and unstructured data, such as booking records, customer reviews, social media interactions, and travel patterns, allowing for personalized marketing strategies and service offerings (Bulchand-Gidumal, William Secin, O'Connor, & Buhalis, 2024). In destination management, big data facilitates demand forecasting, helping tourism operators anticipate peak seasons, adjust pricing strategies, and allocate resources effectively. Insights into traveller behaviour, such as preferred destinations, booking channels, and spending habits, enable targeted campaigns to attract specific customer segments (Hu & Song, 2023). Additionally, real-time data analysis improves operational efficiency by reducing airport wait times, managing visitor flows in popular destinations, and enhancing customer satisfaction through proactive problem-solving. Big data also plays a critical role in sustainability by identifying patterns contributing to over-tourism or underutilization, allowing for balanced resource distribution and policy development. For instance, predictive analytics based on big data can guide eco-tourism initiatives by highlighting destinations that require conservation efforts. Big data transforms tourism into a more adaptive, customer-centric, and sustainable industry by enabling precise and proactive decision-making.

Big data advances sustainability by providing the insights and tools to balance economic growth, environmental preservation, and social well-being. By analyzing extensive datasets from various sources, such as resource usage patterns, environmental impact assessments, and consumer behaviours, big data enables organizations to identify inefficiencies and implement more sustainable practices (Bagherzadeh, Shokouhyar, Jahani, & Sigala, 2021). For example, big data helps monitor visitor flows to prevent overcrowding and preserve natural or cultural heritage sites, ensuring sustainable tourism development. Predictive analytics powered by big data allows businesses to anticipate future demand and optimize resource allocation, reducing waste and minimizing environmental footprints. For instance, energy usage in hotels or transportation networks can be adjusted dynamically based on real-time occupancy and travel data. Big data supports sustainability in agriculture by optimizing water usage, improving crop yields, and reducing chemical inputs through precision farming techniques. Additionally, big data enhances transparency by tracking supply chains and verifying sustainability claims, empowering consumers to make informed choices. Identifying areas of overconsumption or waste guides industries in implementing circular economy models, such as recycling and resource recovery initiatives. On a broader scale, big data supports policymakers in designing effective regulations and strategies by providing evidence-based insights into environmental and societal challenges. Big data drives innovation, fosters accountability, and accelerates the transition toward a more sustainable and resilient global economy through its capacity to deliver precise, actionable information.

6. Conclusions

The research concludes that predictive analytics and advanced machine learning models, such as XGBoost, are transformative in addressing critical challenges in the tourism and hospitality industry. By analyzing 900 data entries containing guest metadata, booking histories, and customer reviews, the XGBoost model achieved a prediction accuracy rate of 85%, demonstrating its effectiveness in forecasting hotel room preferences and occupancy trends. The findings highlight the importance of data-driven decision-making, optimizing resource allocation, personalizing customer experiences, and enhancing operational efficiency. The study further revealed that critical features, such as "Length of Stay" (importance score of 0.35), "Rating"

(0.30), and "Guest Type" (0.25), significantly influenced prediction outcomes, emphasizing the need to focus on these variables in strategic planning. While the model successfully minimized errors in 70% of cases, misclassifications occurred in 12 instances, suggesting the necessity for improved feature engineering and addressing class imbalances. In conclusion, the research underscores the practical value of predictive models in fostering innovation, efficiency, and customer satisfaction within the hospitality sector. These results provide a strong foundation for integrating AI and big data analytics into broader tourism applications, with significant implications for sustainable growth and competitive advantage in the industry.

7. Acknowledgment

The authors sincerely thank the Atma Jaya Catholic University of Indonesia, particularly the Faculty of Business Administration and Communication and Tourism Study Program, for the institutional support in conducting this research. Special thanks to LPPM (Institute for Research and Community Service) and PUSDIPAR (Center of Digital Transformation and Tourism Development) for providing research facilities and guidance throughout this study.

8. Funding

This research received no external funding

9. Conflicts of Interest

The author(s) declare no conflict of interest

References

Al-Qudah, D. A., Al-Zoubi, A. M., Castillo-Valdivieso, P. A., & Faris, H. (2020). Sentiment analysis for e-payment service providers using evolutionary extreme gradient boosting. *IEEE Access*, *8*, 189930–189944. https://doi.org/10.1109/ACCESS.2020.3032216

Al Jassim, R. S., Al Mansoory, S., Jetly, K., & Almaqbali, H. (2024). Enhancing Hotel Performance Prediction in Oman's Tourism Industry: Insights from Machine Learning, Feature Analysis, and Predictive Factors. *IEEE Conference on Evolving and Adaptive Intelligent Systems*. https://doi.org/10.1109/EAIS58494.2024.10570014

Alotaibi, Y., Malik, M. N., Khan, H. H., Batool, A., ul Islam, S., Alsufyani, A., & Alghamdi, S. (2021). Suggestion mining from opinionated text of big social media data. *Computers, Materials and Continua*, *68*(3), 3323–3338. https://doi.org/10.32604/cmc.2021.016727

Ampountolas, A., & Legg, M. (2024). Predicting daily hotel occupancy: a practical application for independent hotels. *Journal of Revenue and Pricing Management*, *23*(3), 197–205. https://doi.org/10.1057/s41272-023-00445-7

Ampountolas, A., & Legg, M. P. (2021). A segmented machine learning modeling approach of social media for predicting occupancy. *International Journal of Contemporary Hospitality Management*, *33*(6), 2001–2021. https://doi.org/10.1108/IJCHM-06-2020-0611

Azhar, A. N., & Khodra, M. L. (2020). Fine-tuning Pretrained Multilingual BERT Model for Indonesian Aspect-based Sentiment Analysis. *2020 7th International Conference on Advanced Informatics: Concepts, Theory and Applications, ICAICTA 2020*. https://doi.org/10.1109/ICAICTA49861.2020.9428882

Bagherzadeh, S., Shokouhyar, S., Jahani, H., & Sigala, M. (2021). A generalizable sentiment analysis method for creating a hotel dictionary: using big data on TripAdvisor hotel reviews. *Journal of Hospitality and Tourism Technology*, *12*(2), 210–238. https://doi.org/10.1108/JHTT-02-2020-0034

Bisoi, S., Roy, M., & Samal, A. (2020). Impact of artificial intelligence in the hospitality industry. *International Journal of Advanced Science and Technology*, *29*(5), 4265–4276.

Bulchand-Gidumal, J., William Secin, E., O'Connor, P., & Buhalis, D. (2024). Artificial intelligence's impact on hospitality and tourism marketing: exploring key themes and addressing challenges. *Current Issues in Tourism*, *27*(14), 2345–2362. https://doi.org/10.1080/13683500.2023.2229480

Chen, M., Xu, H., Wu, Y., & Wu, J. (2024). Sentiment Analysis of Hotel Reviews based on BERT and XGBoost. *Proceedings - 2024 3rd International Conference on Computer Technologies, ICCTech 2024*, pp. 11–15. https://doi.org/10.1109/ICCTech61708.2024.00011

Cheng, Y., Zhao, B., Peng, S., Li, K., Yin, Y., & Zhang, J. (2024). Effects of cultural landscape service features in national forest parks on visitors' sentiments: A nationwide social media-based analysis in China. *Ecosystem Services*, *67*. https://doi.org/10.1016/j.ecoser.2024.101614

Christodoulou, E., Gregoriades, A., Pampaka, M., & Herodotou, H. (2021a). Application of Classification and Word Embedding Techniques to Evaluate Tourists' Hotel-revisit Intention. *International Conference on Enterprise Information Systems, ICEIS - Proceedings*, Vol. 1, pp. 216–223. https://doi.org/10.5220/0010453502160223

Christodoulou, E., Gregoriades, A., Pampaka, M., & Herodotou, H. (2021b). Evaluating the Effect of Weather on Tourist Revisit Intention using Natural Language Processing and Classification Techniques. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, pp. 2479–2484. https://doi.org/10.1109/SMC52423.2021.9658820

Febrian, Y. Y., Wijaya, D. R., & Ervina, E. (2024). Hotel Reservation Cancellation Prediction using Boosting Model. *2024 2nd International Conference on Software Engineering and Information Technology, ICoSEIT 2024*, pp. 138–143. https://doi.org/10.1109/ICoSEIT60086.2024.10497479

Gao, J., Meng, Q., Zhang, L., & Hu, D. (2022). How does the ambient environment respond to the industrial heat island effects? An innovative and comprehensive methodological paradigm for quantifying the varied cooling effects of different landscapes. *GIScience and Remote Sensing*, *59*(1), 1643–1659. https://doi.org/10.1080/15481603.2022.2127463

Herrera, A., Arroyo, Á., Jiménez, A., & Herrero, Á. (2024). Forecasting hotel cancellations through machine learning. *Expert Systems*, *41*(9). https://doi.org/10.1111/exsy.13608

Ho, R. C., Withanage, M. S., & Khong, K. W. (2020). Sentiment drivers of hotel customers: a hybrid approach using unstructured data from online reviews. *Asia-Pacific Journal of Business Administration*, *12*(3–4), 237–250. https://doi.org/10.1108/APJBA-09-2019-0192

Hu, T., & Song, H. (2023). Analysis of Influencing Factors and Distribution Simulation of Budget Hotel Room Pricing Based on Big Data and Machine Learning from a Spatial Perspective. *Sustainability (Switzerland)*, *15*(1). https://doi.org/10.3390/su15010617

Huang, J. (2020). Application of locally linear embedding algorithm on hotel data text classification. *Journal of Physics: Conference Series*, Vol. 1634. https://doi.org/10.1088/1742-6596/1634/1/012014

Kaur, A., Goyal, S., & Batra, N. (2024). Smart Hospitality Review: Using IoT and Machine Learning to Its Most Value in the Hotel Industry. *2024 International Conference on Automation and Computation, AUTOCOM 2024*, pp. 320–324. https://doi.org/10.1109/AUTOCOM60220.2024.10486158

Kozlovskis, K., Liu, Y., Lace, N., & Meng, Y. (2023). Application of Machine Learning Algorithms To Predict Hotel Occupancy. *Journal of Business Economics and Management*, *24*(3), 594–613. https://doi.org/10.3846/jbem.2023.19775

Liu, Z., Jiang, P., Wang, J., Du, Z., Niu, X., & Zhang, L. (2023). Hospitality order cancellation prediction from a profit-driven perspective. *International Journal of Contemporary Hospitality Management*, *35*(6), 2084–2112. https://doi.org/10.1108/IJCHM-06-2022-0737

Mathew, E., & Abdulla, S. (2021). Machine learning to find purchase duration of chain hotels in the UAE. *2021 International Symposium on Networks, Computers and Communications, ISNCC 2021*. https://doi.org/10.1109/ISNCC52172.2021.9615706

Pangestu, M. B., Barakbah, A. R., & Muliawati, T. H. (2020). Data analytics for hotel reviews in multi-language based on factor aggregation of sentiment polarization. *IES 2020 - International Electronics Symposium: The Role of Autonomous and Intelligent Systems for Human Life and Comfort*, 324–331. https://doi.org/10.1109/IES50839.2020.9231625

Reddy, Y. B., Prathima, C., Jaladi, S., Dinesh, B., & Arun Kumar, J. R. (2023). Machine Learning Techniques for Detecting and Analyzing Online Fake Reviews. *Artificial Intelligence and Knowledge Processing: Improved Decision-Making and Prediction*, pp. 271–278. https://doi.org/10.1201/9781003328414-23

Sánchez-Franco, M. J., & Aramendia-Muneta, M. E. (2023). Why do guests stay at Airbnb versus hotels? An empirical analysis of necessary and sufficient conditions. *Journal of Innovation and Knowledge*, *8*(3). https://doi.org/10.1016/j.jik.2023.100380

Satu, M. S., Ahammed, K., & Abedin, M. Z. (2020). Performance Analysis of Machine Learning Techniques to Predict Hotel booking Cancellations in Hospitality Industry. *ICCIT 2020 - 23rd International Conference on Computer and Information Technology, Proceedings*. https://doi.org/10.1109/ICCIT51783.2020.9392648

Serrano, L., Ariza-Montes, A., Nader, M., Sianes, A., & Law, R. (2021). Exploring preferences and sustainable attitudes of Airbnb green users in the review comments and ratings: a text mining approach. *Journal of Sustainable Tourism*, *29*(7), 1134–1152. https://doi.org/10.1080/09669582.2020.1838529

Shakhovska, K., Shakhovska, N., & Veselý, P. (2020). The sentiment analysis model of services providers' feedback. *Electronics (Switzerland)*, *9*(11), 1–15. https://doi.org/10.3390/electronics9111922

Shallan, M. S., Moawad, I. F., El Naggar, R., & Montasser, H. (2024). Using Machine Learning Techniques to Maximize Profitability in the Hospitality Industry. *6th International Conference on Computing and Informatics, ICCI 2024*, pp. 182–188. https://doi.org/10.1109/ICCI61671.2024.10485148

Sreenivas, G., Murthy, K. M., Prit Gopali, K., Eedula, N., & Mamatha, H. R. (2023). Sentiment Analysis of Hotel Reviews - A Comparative Study. *2023 IEEE 8th International Conference for Convergence in Technology, I2CT 2023*. https://doi.org/10.1109/I2CT57861.2023.10126445

Taherkhani, L., Daneshvar, A., Amoozad Khalili, H., & Sanaei, M. R. (2024). Intelligent decision support system using nested ensemble approach for customer churn in the hotel industry. *Journal of Business Analytics*, *7*(2), 83–93. https://doi.org/10.1080/2573234X.2023.2281317

Wang, J., Wu, J., Sun, S., & Wang, S. (2024). The relationship between attribute performance and customer satisfaction: An interpretable machine learning approach. *Data Science and Management*, *7*(3), 164–180. https://doi.org/10.1016/j.dsm.2024.01.003

Yadav, S., Singh, R., Manigandan, E., Unni, M. V., Bhuvaneswari, S., & Girdharwal, N. (2023). Research on Factors Affecting Consumer Purchasing Behaviour on E-commerce Website During COVID-19 Pandemic based on RBF-SVM Network. *2nd International Conference on Automation, Computing and Renewable Systems, ICACRS 2023 - Proceedings*, pp. 371–376. https://doi.org/10.1109/ICACRS58579.2023.10404765

Yoo, M., Singh, A. K., & Loewy, N. (2024). Predicting hotel booking cancelation with machine learning techniques. *Journal of Hospitality and Tourism Technology*, *15*(1), 54–69. https://doi.org/10.1108/JHTT-07-2022-0227

Zhao, M. A., & Jayadi, R. (2021). Forecasting Daily Visitors and Menu Demands in an Indonesian Chain Restaurant using Support Vector Regression Machine. *AIMS 2021 - International Conference on Artificial Intelligence and Mechatronics Systems*. https://doi.org/10.1109/AIMS52415.2021.9466036