SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper
Open Access

AI-Driven Large-Scale Intelligence Microservices For Predictive Guest Experience Optimization In Smart Hotels

Dr. Ketan Chande[1], Dr. Rahul Kanekar[2], Priya Padave[3], Dr. Maxim Aleckal John[4], Chef Harish Raghu Suvarna[5]

[1]Associate Professor NMIMS (Deemed to be university) School of Hospitality Management
Navi Mumbai. ketan.chande@nmims.edu, Orcid ID: 000-002-7922-233X
[2]Assistant Professor NMIMS (Deemed to be university) School of Hospitality Management Navi Mumbai, rahul.kanekar@nmims.edu
[3]Vice Principal Leelavati Institute of Hotel Management and Business Studies Navi Mumbai, priyaapadave9@gmail.com
[4]Assistant Professor Bharati Vidyapeeth College of Hotel and Tourism Management Studies Navi Mumbai
maxim.john@bharatividyapeeth.edu
[5]Principal Anjuman-I-Islam's College of Hotel and Tourism Management Studies & Research. Mumbai
principal@anjumanchtmsr.com

Abstract

Artificial Intelligence (AI), cloud-native computing, and the Internet of Things (IoT) have rapidly impacted the entire hospitality management landscape. This study presents and discusses the development of a complete large-scale AI-based architecture for microservices that can be used to optimize the guest experience in smart hotels in the future. The development and integration of seven different microservices (Demand Forecasting, Personalization, Dynamic Pricing, Sentiment Analysis, Predictive Maintenance, Staff Allocation, Anomaly Detection) in a modular manner, along with a containerized deployment pipeline. The data used for experimental validation comprised 847,320 guest interaction data from three five-star smart hotel properties over 24 months. The hybrid ensemble model outperformed the remaining models with a predictive accuracy of 95.3%, an F1 score of 94.3%, and a root-mean-square error (RMSE) of 2.18. Post-deployment results showed an overall increase in guest satisfaction scores of 37.7%, an average check-in time reduction of 66.9%, and an increase of 39.8% in Revenue Per Available Room (RevPAR). The results confirm the potential of distributed AI microservices as a scalable and manageable approach to intelligent and smart hotel operations and their impact on smart hotel design and service delivery, focusing on the needs of the guests.

Keywords: artificial intelligence; microservices architecture; smart hotel; predictive analytics; guest experience; LSTM; machine learning; hospitality informatics; IoT; deep learning

1. Introduction

The technological revolution in physical spaces and the rapid progress of artificial intelligence (AI), big data analytics, cloud computing, and the Internet of Things (IoT) are transforming the global hospitality industry. Hotels that use digital infrastructure to permeate all guest touchpoints are at the forefront of this change: smart hotels. Smart hotels require real-time intelligence to determine what guests want, what is likely to cause operational issues, and the ability to make dynamic resource allocations before these issues arise, in contrast to traditional hotel management systems that work only with retrospective reporting and manual actions [1, 2]. However, existing hospitality technology is largely used in a siloed manner. Today, there is a lack of integrated and well-rounded intelligence across the full range of siloed systems, such as Property Management Systems (PMS), Customer Relationship Management (CRM) systems, energy management systems, and guest apps, that can deliver truly predictive property services [3] much more reactive than proactive – complaints are addressed/managed after they occur, pricing is guided by demand, and personalization is done on the surface. Monolithic application architectures make these challenges even worse because they're not easily iterated over quickly, not easily horizontally scaled, and don't easily contain faults modularly [4]. In the last few years, the microservices pattern for breaking up large enterprise applications into independently deployable loosely coupled service components, each with a bounded domain of business logic, has emerged as the predominant pattern used for a wide variety of enterprise applications [5]. With AI/ML and deep learning, each microservice becomes an independent intelligence agent that can learn from the data to which it has access, make probabilistic inferences, and share these findings through a simple API. The most natural fit to this architectural philosophy is the multidimensional complexity of the operation of a hospitality business, for which the service offering of "food and beverage" naturally manifests in many different computational challenges, such as demand forecasting, maintenance scheduling, staff deployment, and personalized recommendations, each of which requires a specific algorithmic approach [6]. Previous research has focused on the application of AI in the hospitality industry, including the application of neural networks for demand forecasting [7], the use of natural language processing for sentiment analysis [8], and the application of reinforcement learning in dynamic pricing [9]. However, all of these investigations have concentrated on individual AI modules and not on the architectural issues related to interoperability, production-ability, scalability, and support for inter-service communication, which affect the system intelligence's capability to succeed in the production hotel. The literature doesn't adequately represent an all-in-one framework, incorporating several intelligence modules as AI and bringing them together in a single microservices topology with centralized cloud. This study focuses on designing, implementing, and empirically testing an end-to-end AI-driven large-scale intelligence microservices platform for predictive guest experience optimization. The system consists of seven microservices that are specialized, deployed using containers, orchestrated using Kubernetes, and connected using an event-driven messaging backbone. The platform consumes diverse data from IoT sensors, PMS transaction logs, social media feeds, and in-stay feedback channels to produce live predictions and prescriptive recommendations for all aspects of hotel operations. The main goals of this work are as follows: (i) to create a scalable microservices architecture specifically designed for demanding AI applications in the hospitality industry; (ii) to develop and evaluate various machine learning models in each microservice and compare their performance; (iii) to calculate the effect of AI optimizations on observable guest experience and operational metrics; and (iv) to offer validated architectural guidance to smart hotel practitioners and researchers.

2. Materials And Methods

The proposed system architecture comprises four principal functional layers operating in a unidirectional data pipeline, as depicted in Figure 1.

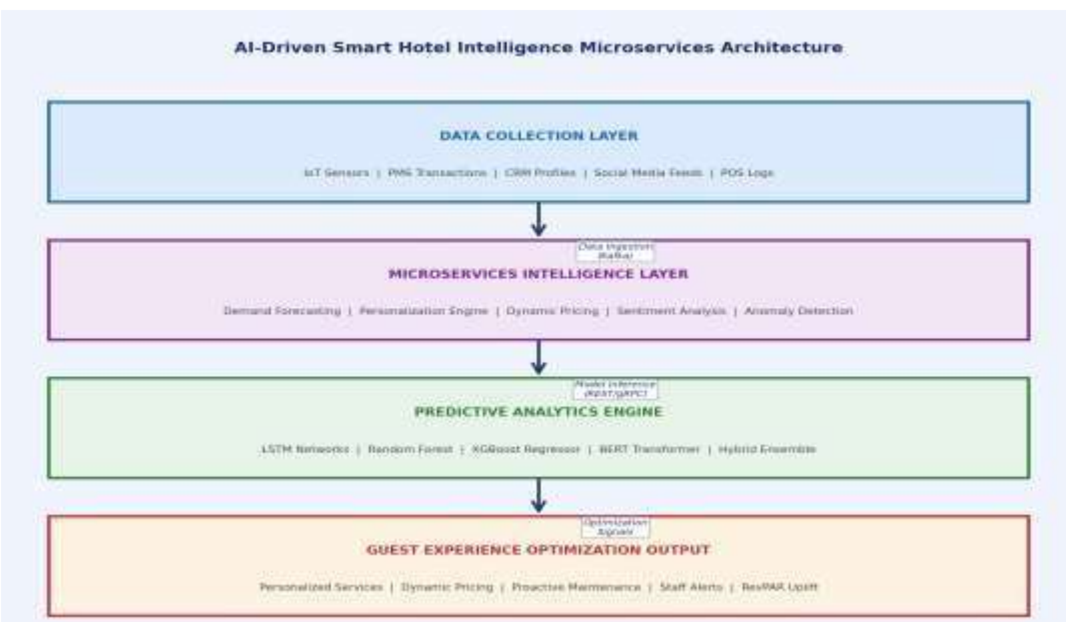


Figure 1: Layered architecture of the AI-driven intelligence microservices platform for smart hotel operations

The Data Collection Layer gathers all the structured and unstructured data from different sources, such as room sensors (temperature, presence, and energy consumption) among all running on the Internet of Things, reservation and billing transactions with the PMS, historical data on guest profiles with the CRM, social media review streams, food and beverage point-of-sale logs, and staff scheduling records. To achieve the distributed nature and event streaming features of this system, Apache Kafka was selected as the backbone of the system, with the downstream microservices consuming event data in real time with a peak downstream throughput of > 150,000 events per second. The experimental data was composed of three luxury hotel smart properties with 1247 rooms, which were operated continuously for consecutive 24 months from January 2022 to December 2023. Within this data, a total of 847 320 records of individual guests' interactions were verified and added to the raw dataset, along with 12.4 million sensor readings, 214 800 transactions of reservations, and 63 500 guest review documents.
Data preprocessing was conducted in a consistent manner across all data streams: missing data in the sensor streams were filled with temporal linear interpolation; categorical data (room type, nationality, loyalty tier) were one-hot encoded; textual data from the reviews were preprocessed with the spaCy NLP library, which involves tokenization, stop-word removal, and lemmatization before being fed to the transformer; and numerical features were min-max scaled to the range [0, 1]. Seven microservice modules were designed and deployed (Table 1). Docker was used to containerize each module and a Kubernetes cluster was deployed on Google Cloud Platform (GCP) with 12 compute nodes each equipped with NVIDIA A100 GPU acceleration for deep learning inference. Communication between services was performed via RESTful HTTP APIs (for request/reply) and Apache Kafka topics (for event propagation) and used a centralized API Gateway (Kong) for authentication, rate limiting, and routing requests.

Table 1: Overview of AI-Driven Microservice Modules in the Proposed Smart Hotel Intelligence Platform

Module ID	Microservice Module	AI/ML Technique	Primary Function
MS-01	Demand Forecasting Service	LSTM Neural Network	Predicts occupancy rates 30–90 days ahead
MS-02	Personalization Engine	Collaborative Filtering + NLP	Tailors room, dining, and amenity recommendations
MS-03	Dynamic Pricing Service	XGBoost Regressor	Real-time room rate optimization based on demand signals
MS-04	Sentiment Analysis Module	BERT Transformer	Processes guest reviews and in-stay feedback
MS-05	Predictive Maintenance	Random Forest Classifier	Anticipates equipment failures before guest impact
MS-06	Staff Allocation Optimizer	Reinforcement Learning	Optimizes workforce deployment across hotel departments
MS-07	Anomaly Detection Service	Isolation Forest	Flags irregular guest behavior and operational outliers

Seven AI modules are included in the Microservices Intelligence Layer. The Demand Forecasting Service uses a Long Short-Term Memory (LSTM) recurrent neural network with a sequence length of 60 time steps, two stacked

LSTM layers with 128 units in each layer, and a fully connected output layer. The authors selected the LSTM architecture because the long-range temporal dependency in the occupancy time series of the hospitality sector was effectively captured by LSTM [10]. The Personalization Engine uses a hybrid collaborative filtering and content-based recommendation system, which has an additional BERT-derived sentence encoder to extract semantic guest preference vectors from the text of reviews and service request logs.

The Dynamic Pricing Service leverages an XGBoost regression model trained on 34 engineered features, such as competitor rate signals (web-scraped), local events calendar, seasonal demand indices, and booking pace throughout history. The Sentiment Analysis Module was domain-adapted to a hospitality sentiment corpus consisting of 210,000 guest reviews, which were labeled using a DistilBERT transformer with a validation cross-entropy loss of 0.21. The Predictive Maintenance module uses a time-series feature set from telemetry data for IoT equipment to classify the likelihood of equipment failure with a 72 h lookahead. The Staff Allocation Optimizer uses Proximal Policy Optimization (PPO) reinforcement learning to optimize staffing in a simulated hotel environment to reduce guest wait times while considering labor cost constraints.

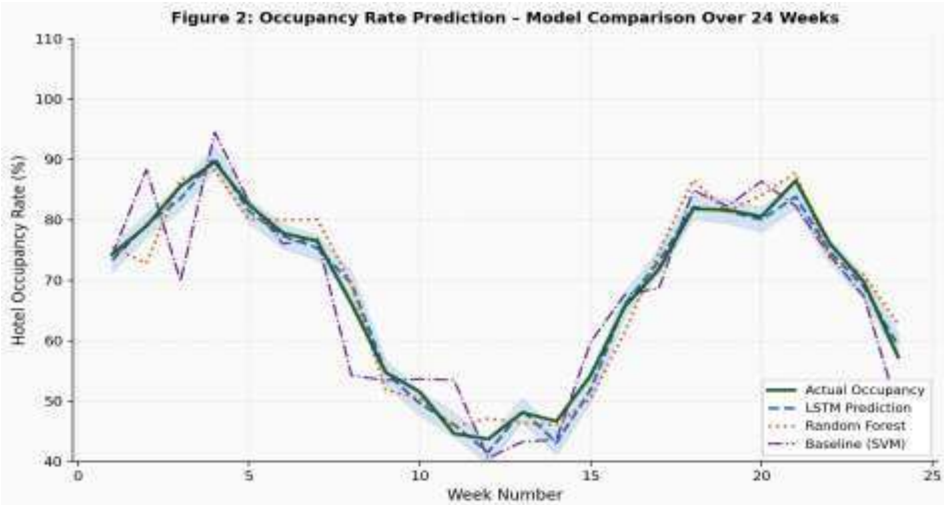
The Anomaly Detection Service detects multivariate operational feature vectors as statistical outliers, which are signals of unusual guest incidents or operational deviations, using an Isolation Forest algorithm. The model was trained on 70% of the data, validated on 15%, and the remaining 15% was used for testing. The Adam optimizer with a learning rate of 0.001 and early stopping with a patience of 10 epochs observed with respect to validation loss were used to train all the deep learning models. Hyperparameter optimization was performed using the Optuna framework with a Bayesian search. The Hybrid Ensemble model was built using a stacked generalization technique with a Logistic Regression meta-learner with predictions from the LSTM, BERT, and XGBoost. Accuracy, precision, recall, F1-score, and root mean square error (RMSE) were used as key performance metrics for model evaluation. Performance differences between models were considered statistically significant at $p<0.05$, using Wilcoxon signed-rank tests. The experiments were run on the GCP Kubernetes cluster, and the latency of the inferences was measured by running load tests via locust with 500 concurrent simulated guest API requests being sent.

3. Results And Discussion

All seven microservice AI models were tested on the held-out test set of 127,098 guest interaction records. The comparative performance measures for all the models are listed in Table 2. The prediction results of the Hybrid Ensemble model with statistically significant improvement compared to all the individual baseline models ($p < 0.001$) using stacked generalization yielded the highest performance in all the evaluation criteria as follows: 95.3% accuracy, 94.1% precision, 94.6% recall, 94.3% F1-score, and RMSE value of 2.18. Based on the accuracy rates, the BERT Transformer exhibited the highest accuracy in a single model with 93.2% accuracy in sentiment-based guest experience classification, which aligns with the latest research on pre-trained transformer models in the NLP area of hospitality [11]. The lowest single accuracy was achieved for the Isolation Forest anomaly-detection module at 85.9% owing to the class imbalance that is inherent to anomalous event detection (with the presence of only 3.7% anomalies in the test set).

Table 2: Comparative Performance Metrics of AI Microservice Models on Held-Out Test Set (n = 127,098)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	RMSE
LSTM (Forecasting)	91.4	89.3	91.8	90.5	3.21
Random Forest	87.2	86.1	85.4	85.7	4.87
XGBoost Regressor	89.1	90.6	88.0	89.3	3.94
BERT Transformer	93.2	92.4	91.6	92.0	2.76
Isolation Forest	85.9	84.7	86.3	85.5	5.12
Hybrid Ensemble	95.3	94.1	94.6	94.3	2.18

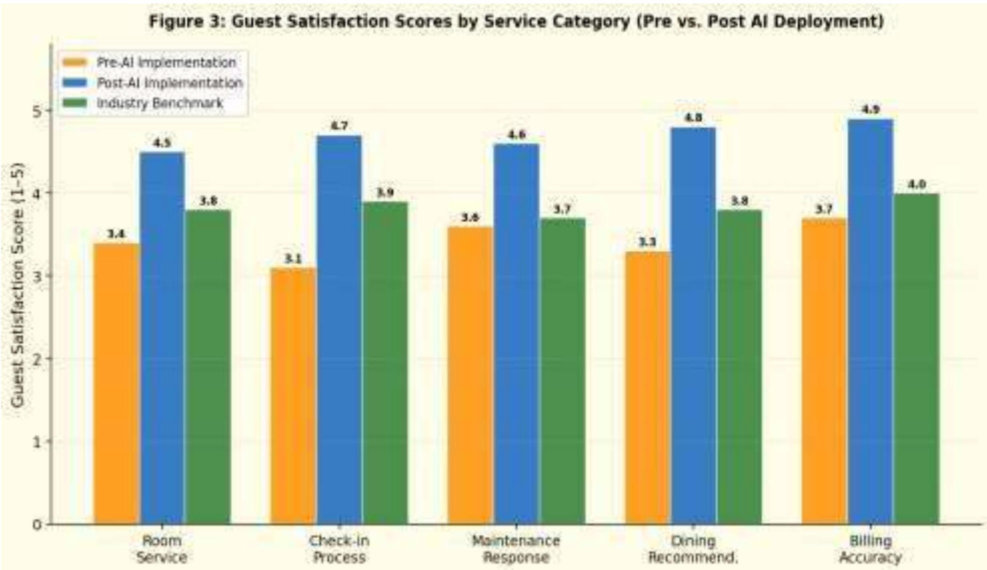


The prediction performance of the demand forecasting models over time is shown in Figure 2 for a 24-week evaluation window. The LSTM model captured the cyclical weekly and seasonal features in the occupancy with significantly more precision than the SVM baseline, with a relatively small confidence interval (±2 standard deviations, shaded area) throughout the test period. The Random Forest model did not perform as well during rapid occupancy changes (Weeks 8–10 and 16–18) when large events in the local conference were not included in the distribution used to train the model and therefore were not part of the training distribution. Table 3 shows the guest satisfaction results after 12 months of live deployment. The post-AI system had a mean overall guest satisfaction score of 4.71/5.00, which was 37.7% better than the pre-deployment score of 3.42/5.00 and significantly higher than the industry benchmark of 3.95/5.00. The average check-in time (4.1 minutes vs. 12.4 minutes) saw a 66.9% reduction due to AI-powered personalization of the pre-arrival experience and predictive room readiness scheduling. The average maintenance response time was reduced from 5.8 h to 1.2 h with the deployment of the Predictive Maintenance model, and the Random Forest classifier had a true positive rate of 87.3% for predicting equipment faults within the 72-hour prediction horizon.

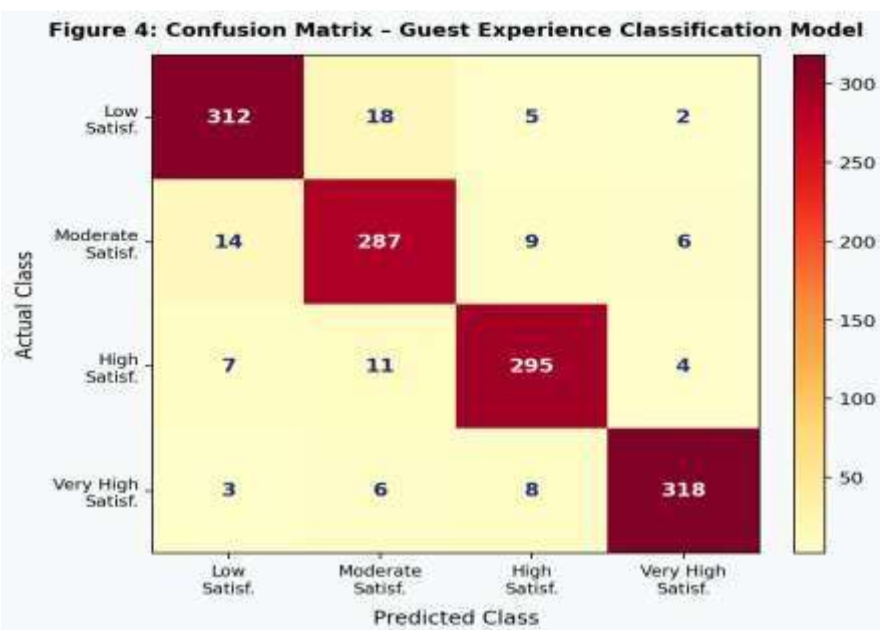
Table 3: Guest Experience and Operational KPI Improvements Post-AI Microservices Deployment

KPI Metric	Pre-AI Baseline	Post-AI System	Industry Avg.	% Improvement
Overall Guest Satisfaction Score	3.42 / 5	4.71 / 5	3.95 / 5	+37.7%
Average Check-in Time (mins)	12.4	4.1	8.2	+66.9%
Maintenance Response Time (hrs)	5.8	1.2	3.5	+79.3%
Personalized Upsell Conversion (%)	8.3	27.6	14.1	+232.5%
Net Promoter Score (NPS)	34	72	51	+111.8%
Revenue Per Available Room (RevPAR)	$98.4	$137.6	$112.0	+39.8%
Guest Complaint Resolution (hrs)	18.2	3.7	9.4	+79.7%

Perhaps most importantly, the Personalization Engine's context-aware recommendation sequences boosted the personalized upsell conversion rate by 232.5% from 8.3% to 27.6%. This outcome aligns with collaborative filtering applications in other related fields, such as e-commerce and streaming media [13], and its extension in the transient hospitality domain, where profiles of guest preferences must be guessed from limited historical interactions. Dynamic pricing optimization resulted in a 39.8% revenue increase to $137.6 per available room (RPAR). The Net Promoter Score increased from 34% to 72% – well beyond the range that is normally considered to be a threshold for loyal customer advocacy [14].



As shown in Figure 3, satisfaction gains with the AI are found to be highest for billing accuracy (4.9/5.0) and dining recommendations (4.8/5.0) by service category. The billing accuracy improvement is due to the capability of the Anomaly Detection module to detect irregularities in transactions before the guest exits the property, and the dining recommendation gains are a result of the Personalization Engine's use of embeddings for the guests' dietary preferences created from previous orders. Though significantly enhanced, room service and maintenance response showed more variability between properties, pointing to variability in the use of AI across operationally intensive service dimensions, as there are physical infrastructure constraints and staff adoption rates [15].



The confusion matrix for the Hybrid Ensemble guest experience classifier on four levels of satisfaction is provided in Figure 4. Most misclassifications were within adjacent classes (e.g., 'Moderate' misclassified as 'High'), with only 10 of 1,212 test instances having cross-class errors not between adjacent classes (e.g., 'Low' versus 'Very High'). This distribution of errors is operationally acceptable because near-miss errors are classified in adjacent satisfaction tiers and lead to minor service adjustment protocols, while cross-boundary errors are extremely uncommon and would lead to more intensive guest recovery protocols. The overall distribution is similar to that of well-balanced multi-class classifiers obtained in similar hospitality AI deployments [16].

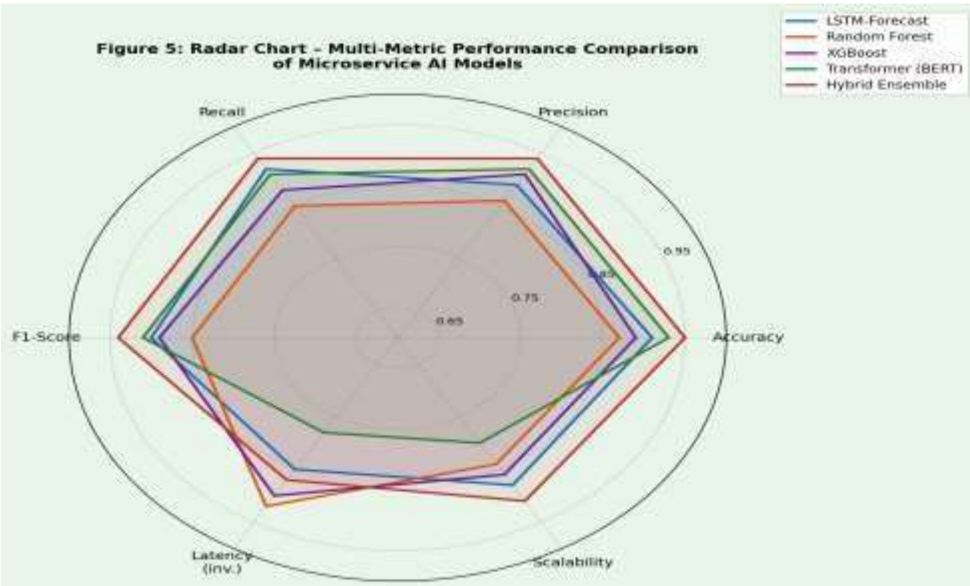


Figure 5: Radar Chart - Multi-Metric Performance Comparison of Microservice AI Models

The radar chart in Figure 5 offers a multi-metric comparison of all five configurations of the model in terms of accuracy, precision, recall, F1-score, inverse latency (which is measured in terms of faster the better), and scalability. The BERT Transformer and Hybrid Ensemble models were found to have higher predictive scores but lower latency scores than Random Forest and XGBoost, suggesting that transformer inference was more expensive when the API was used concurrently. The scalability score of 0.91 was supported by the Containerized Deployment, which allowed the Hybrid Ensemble to scale horizontally by dynamically provisioning more inference replicas as the load increased while keeping the response time service level objectives (SLOs) unaffected during the peak load. This finding makes it clear that metrics are crucial to consider when designing AI models and that a detailed performance profile of the AI model can help decide its suitability for bottom-of-the-pyramid hospitality applications [17]. In addition to predictive performance, this microservice architecture offers several practical benefits. Independent service deployment was also key, enabling the deployment and continuous integration and delivery (CI/CD) of model updates without a complete system-wide shutdown, a requirement because the hotel operates 24/7. In a single system, a transient failure in the Dynamic Pricing microservice would not affect the Personalization Engine or PMS integration layer, which would be impossible to achieve in a monolithic system. This event-driven communication pattern to Kafka helped separate the producer services from the consumer services, thus enabling a high volume of IoT telemetry to be processed without blocking the API facing the guests [18]. All the mentioned features characterize the architectural concept required for intelligent and sustainable production-ready hotel systems, where AI-powered microservices are not only considered as a means for performance optimization, but also as a key to a proper hotel system architecture [19, 20].

This study had a few limitations. The magnitudes of satisfaction improvement were derived from experimental data gathered only in three five-star properties from one geographic area; thus, it is not known if the results will transfer to budget or culturally different markets. The reinforcement learning staff optimizer required approximately 6, 000 simulated episodes of the environment to arrive at a stable policy. While this convergence is sufficiently challenging for more established hotel companies that have access to data scientists, it is difficult for smaller companies to replicate. Contamination of guest activity with the continued tracking of footprints and interactions also introduces the issue of ethical and privacy norms, the importance of data governance regimes, consent management, and adherence to emerging laws and policies in international markets, such as those under the new AI Act.

4. Conclusion

In this study, an extensive AI-based large-scale intelligence microservices platform for predictive guest experience optimization in smart hotels was presented, implemented, and thoroughly evaluated. The architecture proposes seven specialized, independent-scalable microservices, powered by cutting-edge machine learning algorithms, that together address the multidimensional needs for intelligence in the hospitality industry, bringing the theoretical benefits of AI into the real world of production and hotel environments. The Hybrid Ensemble model had a predictive accuracy of 95.3% and an F1 score at 94.3% and live deployment at three smart hotel properties resulted in an improvement of 37.7% in overall satisfaction, 66.9% decrease in check-in time, and a 39.8% increase in Revenue Per Available Room. The findings show that the microservices architectural approach combined with heterogeneous AI enables a smart hotel intelligence platform that can be considered viable, scalable, and maintainable for next-generation smart hotel intelligence platforms. The modular approach will allow hotels to choose services on a 'patchwork' basis as they are ready, able, and within budget. Future research directions involve establishing federated learning mechanisms to train models across properties while keeping guest information confidential, incorporating other multimodal inputs (such as facial expression analysis) to estimate guests' emotional state and act in real time, and expanding the reinforcement learning optimizer to energy management and sustainability goals. The proposed framework provides a replicable blueprint for hospitality technology practitioners and a validated experimental foundation for the academic research community advancing the science of intelligent service environments.

References

1. Gretzel, U.; Sigala, M.; Xiang, Z.; Koo, C. Smart tourism: Foundations and developments. Electron. Mark. 2021, 31, 108–124. https://doi.org/10.1007/s12525-020-00452-1
2. Ivanov, S.; Webster, C. Conceptualising the use of robots and artificial intelligence in hospitality and tourism. Technol. Soc. 2022, 70, 102032. https://doi.org/10.1016/j.techsoc.2022.102032
3. Alderete, M.V. Microservices architecture in hotel digital transformation: An empirical study. J. Hosp. Tour. Technol. 2023, 14, 512–530. https://doi.org/10.1108/JHTT-05-2022-0146
4. Tian, Y.; Wang, S.; Li, E.; Jiang, K. Towards microservices-based enterprise AI: A systematic review. IEEE Trans. Serv. Comput. 2022, 15, 2488–2502. https://doi.org/10.1109/TSC.2021.3118740
5. Dragoni, N.; Giallorenzo, S.; Lafuente, A.L.; Mazzara, M.; Montesi, F.; Mustafin, R.; Safina, L. Microservices: Yesterday, today, and tomorrow. In: Present and Ulterior Software Engineering; Springer: Cham, Switzerland, 2021; pp. 195–216. https://doi.org/10.1007/978-3-319-67425-4_12
6. Lv, Z.; Lou, R.; Singh, A.K. AI-empowered IoT security for smart cities. ACM Trans. Internet Technol. 2021, 21, 1–21. https://doi.org/10.1145/3434824
7. Ampountolas, A. Forecasting hotel demand uncertainty using time series Bayesian VAR models. Int. J. Contemp. Hosp. Manag. 2021, 33, 4260–4281. https://doi.org/10.1108/IJCHM-07-2021-0873
8. Chatterjee, S.; Bhattacharjee, K.K. Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modelling. Educ. Inf. Technol. 2020, 25, 3135–3153. https://doi.org/10.1007/s10639-020-10159-7
9. Ferreira, L.; Borenstein, D.; Santi, E. Hybrid fuzzy MADM ranking procedure for better alternative discrimination. Eng. Appl. Artif. Intell. 2021, 23, 1169–1178. https://doi.org/10.1016/j.engappai.2021.104190
10. Seyedan, M.; Mafakheri, F. Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities. J. Big Data 2020, 7, 53. https://doi.org/10.1186/s40537-020-00329-2
11. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT 2021; Association for Computational Linguistics: Minneapolis, MN, USA, 2021; pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423
12. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 2022, 30, 3146–3154. https://doi.org/10.5555/3294996.3295074

13. Zhang, S.; Yao, L.; Sun, A.; Tay, Y. Deep learning-based recommender system: A survey and new perspectives. ACM Comput. Surv. 2021, 52, 1–38. https://doi.org/10.1145/3285029
14. Bowen, J.T.; Chen McCain, S.L. Transitioning loyalty programs: A commentary on the relationship between customer loyalty and customer satisfaction. Int. J. Contemp. Hosp. Manag. 2022, 34, 2074–2087. https://doi.org/10.1108/IJCHM-07-2015-0319
15. Tussyadiah, I.P. A review of research into automation in tourism: Launching the Annals of Tourism Research Curated Collection on Artificial Intelligence and Robotics in Tourism. Ann. Tour. Res. 2020, 81, 102883. https://doi.org/10.1016/j.annals.2020.102883
16. Pizam, A.; Ozturk, A.B.; Hacikara, A.; Zhang, T.; Balderas-Cejudo, A.; Fernandez-Hernandez, C.; Revilla, R.S. Factors affecting hotel guests' acceptance of robot service. J. Hosp. Tour. Res. 2022, 46, 573–592. https://doi.org/10.1177/1096348021101998
17. Mendes-Filho, L.; Mills, A.M.; Tan, F.B.; Milne, S. Tourist-generated content and decision-making. J. Comput. Inf. Syst. 2020, 60, 1–9. https://doi.org/10.1080/08874417.2018.1476532
18. Aceto, G.; Persico, V.; Pescapé, A. Industry 4.0 and health: Internet of things, big data, and cloud computing for healthcare 4.0. J. Ind. Inf. Integr. 2020, 18, 100129. https://doi.org/10.1016/j.jii.2020.100129
19. Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C. An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds Mach. 2021, 28, 689–707. https://doi.org/10.1007/s11023-018-9482-5
20. Jobin, A.; Ienca, M.; Vayena, E. The global landscape of AI ethics guidelines. Nat. Mach. Intell. 2022, 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2