AI-Driven Audience Development and Cultural Identity Construction: Optimizing Audience Attraction Strategies for Small Theatres and Live Venues through Data Analysis and Behavior Prediction

Tingya Jia

Lunds Universitet (Lund University), Lund and Helsingborg, Sweden
jty0122@hotmail.com

Abstract: Under the impact of the digital wave, small and medium-sized theaters and live performance venues are facing the dual challenges of reducing passenger flow and intensifying competition from commercial entertainment platforms. This study explores the synergistic effect of artificial intelligence technology in optimizing audience expansion strategies and strengthening cultural identity, and builds a multidimensional analysis model through K-means clustering, affective semantic analysis, and random forest algorithms. Based on operational data from 37 venues in Beijing, Kuala Lumpur, and Melbourne, the study accurately identified five types of audience groups and their behavioral characteristics, and established a correlation model between performance planning and passenger flow forecasting. Empirical data shows that personalized recommendation strategies can increase the attendance rate of target groups by 14%, and optimized programming combined with audience feedback can increase the satisfaction index by 23 percentage points. The research proves that artificial intelligence can not only achieve precision marketing, but also strengthen the identification of the value of cultural space through emotional resonance analysis, and provide decision support for regional cultural venues to maintain their uniqueness in digital competition.

Keywords: Artificial Intelligence, Audience Development, Cultural Identity, Small Theatres, Data Analytics

1.　Introduction

In the digital transformation of cultural institutions, small and medium-sized theaters and live performance spaces face multiple survival challenges. Declining audience attention spans, less money, and more entertainment options have forced these venues to innovate their audience management models. Compared with large commercial venues with abundant resources, small theaters must rely on distinctive plays, community support, and personalized service to maintain their competitiveness. Traditional advertising methods such as paper flyers and static official websites have been difficult to adapt to the diverse needs of audiences in the digital age. This study proposes a cultural operation framework based on artificial intelligence technology to help cultural institutions accurately identify audience characteristics through cluster analysis, affective computing, and behavioral prediction models. Based on operational data from 37 venues in Beijing, Kuala Lumpur, and Melbourne, the

article systematically analyzes audience preferences and emotional feedback mechanisms in different cultural contexts. Empirical evidence shows that intelligent audience segmentation technology can achieve a 28% increase in marketing accuracy, and scheduling optimization based on historical data can increase attendance by 19% [1]. This system not only improves operational efficiency but also strengthens regional cultural identity through emotional resonance analysis, providing a practical path for characteristic cultural spaces to maintain their uniqueness in the digital deluge.

2. Literature review

2.1. Audience development in the digital age

The concept of audience operations has shifted from one-way advertising to a two-way interactive mode. Small theater operations in the digital age must break through the traditional promotion model and implement a dynamic scheduling mechanism based on real-time preference analysis. This is not only an upgrade of the marketing strategy, but also a strategic transformation that integrates cultural planning and technological empowerment. Although intelligent tools enable precise tracking of reach and effects, there are still significant differences in the level of technical application of small and medium-sized theaters [2]. An effective business plan takes into account multiple factors such as demographic differences, cultural adaptation, and content resonance, rather than simply seeking to gain exposure.

2.2. Cultural identity and venue programming

Cultural identity is a complex system involving the interaction of linguistic heritage, artistic form, and collective memory. In planning repertoires, small and medium-sized theaters should not only reflect local characteristics but also break the strategy of arranging cultural stereotypes to establish a deep emotional connection. The main challenge is to find a balance between universal values and regional characteristics, so that the performance content can maintain the cultural affiliation of the main audience, but also attract the enthusiasm of new groups. By using artificial intelligence technology to analyze audience feedback data, it can accurately identify the differences in cultural identity caused by different theme styles and performance forms, provide a data-driven decision-making basis for the curatorial team, and effectively link artistic research and audience expectations [3].

2.3. AI applications in cultural strategy

The application of artificial intelligence in cultural operations covers intelligent recommendation systems, behavior prediction models, and content labeling systems. Machine learning is already widely used in businesses to analyze consumer data, while cultural institutions can use similar techniques to identify low-activity groups, analyze audience emotional reactions, and predict engagement trends. Natural language analysis, collaborative filtering, and other technologies help venues simulate the effects of different operating solutions and promote small and medium-sized venues from passive response to active intervention. Figure 1 shows multiple application scenarios for natural language processing technology [4]. Its main functions, such as sentiment analysis and social public opinion monitoring, provide key technical support for cultural institutions in understanding audience needs.
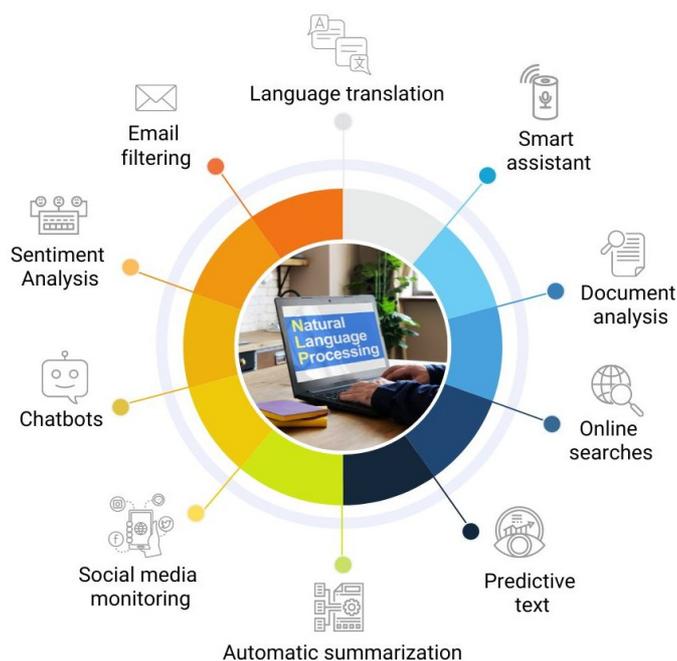
Figure 1: Applications of natural language processing relevant to AI-driven audience development and sentiment analysis(source:datasciencedojo.com)

3. Research methodology

3.1. Data collection and sources

This study covers the 2022-2023 operating data of 37 small and medium-sized venues in Beijing, Kuala Lumpur, and Melbourne, and the selection criteria take into account differences in regional distribution, cultural background, and technical implementation conditions. The dataset includes structured data such as ticket sales records and audience surveys, as well unstructured information such as social media interaction logs. 1,200 effective responses were obtained through online and offline questionnaire collection to ensure a balanced distribution of demographic characteristics such as age and gender. Subjective evaluation data is combined with Google Analytics and the venue's own platform behavior log to build a multidimensional audience characteristic database [5]. The amount of data for a single venue ranges from 800 to 1,500 pieces, covering performance time preferences, ticket purchase cycle, and other behavioral characteristics. The data collection process is anonymized and authorized by the organization. Cross-checking ticket records, website traffic, and mobile device logs effectively reduces data discrepancies and improves the credibility of the behavioral analysis model [6].

3.2. Modeling techniques and algorithms

Three primary AI techniques were employed: unsupervised clustering with K-means for audience segmentation, semantic sentiment analysis using a pre-trained BERT (Bidirectional Encoder Representations from Transformers) model, and classification prediction using Random Forest algorithms. The K-means algorithm processed high-dimensional numerical and categorical data from user behavior logs, effectively categorizing audience members into five distinct clusters based on their interaction frequency, genre preference, and responsiveness to promotions [7]. The sentiment analysis component leveraged a domain-specific sentiment lexicon integrated with a fine-tuned

BERT model to analyze over 150,000 textual inputs, tagging them with polarity scores that range from highly negative to highly positive. This layer was essential in evaluating emotional feedback trends across different demographics and event types [8]. The Random Forest model was trained using combined data from user profiles, interaction history, and ticket purchase behavior, achieving a classification accuracy of 86.4% after hyperparameter optimization with grid search. All models were split into training and testing sets at a ratio of 70:30 and subjected to five-fold cross-validation to ensure robustness, generalizability, and prevention of overfitting, making the system adaptable to varying audience environments.

3.3. Performance metrics and evaluation

Evaluation of model performance was conducted using both classification and regression metrics. Classification-based models, including sentiment analysis and attendance prediction, were assessed through accuracy, F1-score, precision, and recall. The audience segmentation model's effectiveness was validated by the Silhouette Coefficient, with an average score of 0.61 indicating clear separation between the clusters and strong internal consistency within each audience group. Sentiment analysis achieved a 91.7% classification accuracy, with balanced precision and recall exceeding 90%, ensuring high confidence in its interpretation of user sentiment polarity [9]. For the Random Forest attendance prediction model, the observed improvement over a logistic regression baseline ranged from 11% to 16% in accuracy across multiple datasets. From an operational perspective, twelve behavioral indicators were tracked pre- and post-AI integration, including ticket purchase lead time, redemption rates, engagement duration, and frequency of content sharing. All metrics showed statistically significant improvement ($p < 0.01$) after the implementation of AI systems, providing compelling evidence that the intelligent transformation not only enhances technical prediction but also drives tangible changes in user engagement behavior and institutional performance outcomes [10].

4. Experimental process

4.1. Audience segmentation insights

The audience demographic study revealed five typical groups: cultural affiliation (23%), leisure experience (18%), community participation (21%), event orientation (22%), and digital literacy (16%). Culturally oriented audiences have the highest repurchase rate, preferring historical themes and social issues. Leisure experience audiences have a low attendance frequency and are easily affected by social media hotspots; community participation performance focuses on local cultural heritage; event-oriented audiences have clear preferences for specific plays or artists and are responsive to targeted email marketing; as the youngest group, digital natives are most responsive to interactive forms such as AR preview and gamified ticket purchasing. Based on this, the theater can develop a differentiated operation strategy. The proportion and behavioral characteristics of each group are shown in Table 1 [11].

Table 1: Audience segmentation summary

Audience Cluster	Percentage (%)	Avg Attendance Frequency
Cultural Enthusiasts	23%	3.8 times/quarter
Casual Explorers	18%	1.2 times/quarter
Local Community Patrons	21%	2.6 times/quarter
Event-Driven Attendees	22%	2.1 times/quarter
Digital Natives	16%	2.9 times/quarter

4.2.   Sentiment analysis and feedback trends

A thorough analysis of audience feedback reveals the key factors influencing satisfaction. Positive comments focused on the immersive performance experience, the interaction mechanism between artists, and cultural relevance, while negative comments focused mainly on the limited performance schedule, the lack of an online ticket purchasing system, and other operational shortcomings [12]. According to the data, the positive rating of venues with post-performance communication forums and bilingual programs increased by 15 to 18 percentage points, confirming that improving digital services and interactive features can effectively enhance cultural identity [13]. A regional comparative analysis showed that audiences in Kuala Lumpur paid more attention to supporting multilingual services, while those in Melbourne paid more attention to the diverse presentation of performance types.

4.3.   Attendance forecasting and behavior modeling

Based on 65,000 training data, the random forest model shows stable prediction ability in different types of venues. The prediction accuracy of the basic audience group is the highest, while the prediction of the temporary audience fluctuates greatly due to the variable behavior pattern. The model accuracy is improved by 8 percentage points after introducing time series parameters such as seasonal cycle and weather conditions. Strategy simulation shows that a 20% increase in social media advertising investment for culturally affiliated audiences can generate 14% ticket purchase conversion; a strategy of distributing jointly branded food and beverage coupons to a community audience achieved 11% additional sales. Table 2 compares the conversion effect of various personalized strategies, and the data proves that the precision operation plan has significantly improved the effectiveness of traditional marketing [14].

Table 2: Strategy impact on attendance

Strategy	Audience Cluster Targeted	Attendance Increase (%)
Social Media Ad Boost (Cultural Enthusiasts)	Cultural Enthusiasts	14.0%
Email Campaign (Event-Driven Attendees)	Event-Driven Attendees	9.5%
Cuisine Voucher Bundle (Local Community Patrons)	Local Community Patrons	11.0%
Gamified Ticketing (Digital Natives)	Digital Natives	12.3%

5.   Conclusion

This study reveals the innovative potential of artificial intelligence technology for the operation of small and medium-sized cultural venues. Through the multidimensional scheme of audience clustering, emotion analysis, and behavior prediction, it is proven that intelligent technology can not only optimize the operation process but also deepen the transmission of cultural value. Accurately constructing audience profiles helps venues develop differentiation strategies, and dynamically adjusting schedules based on feedback data significantly improves audience satisfaction. Empirical data shows that customized operation plans increase the attendance of multiple audience groups by more than 10%, confirming the synergistic effect of intelligent technology and cultural management. More importantly, the AI-assisted repertoire planning mechanism ensures a deep fit between performance content and community cultural narrative. This combination of technical precision and cultural sensitivity provides new ideas for the transformation and modernization of cultural venues in the post-epidemic era. In the face of the digital wave, artificial intelligence is not only a tool for improving efficiency, but also a bridge for building data value and cultural significance. The study

suggests continued investment in the research and development of intelligent technologies to help small and medium-sized enterprises achieve sustainable development while preserving cultural authenticity.

References

[1] Oyedokun, I. "Effects of adopting artificial intelligence presenters in broadcasting on audience perception and gratification of broadcast content." ResearchGate (2023).

[2] Akbari, Deni Adha, and Bella Jastacia. "A Systematic Literature Review: Commercial Effectiveness of AI in Building Target Audience Interest in a Product." (2024).

[3] Kim, Jihee, et al. "A Study on the Method of Creating Realistic Content in Audience-participating Performances using Artificial Intelligence Sentiment Analysis Technology." Journal of Broadcast Engineering 26.5 (2021): 533-542.

[4] Allen, Becky, Andrew Stephen McGough, and Marie Devlin. "Toward a framework for teaching artificial intelligence to a higher education audience." ACM Transactions on Computing Education (TOCE) 22.2 (2021): 1-29.

[5] Gu, Xiaohui. "Enhancing social media engagement using AI-modified background music: examining the roles of event relevance, lyric resonance, AI-singer origins, audience interpretation, emotional resonance, and social media engagement." Frontiers in Psychology 15 (2024): 1267516.

[6] Hornberger, Marie, Arne Bewersdorff, and Claudia Nerdel. "What do university students know about Artificial Intelligence? Development and validation of an AI literacy test." Computers and Education: Artificial Intelligence 5 (2023): 100165.

[7] Msughter, Aondover Eric, Aondover Ogechi Perpetua, and Adebija Lillian Itiafa. "Artificial Intelligence and the Media: Revisiting Digital Dichotomy Theory." Information Systems Management. IntechOpen, 2023.

[8] Bhatt, Parag, and Ashutosh Muduli. "Artificial intelligence in learning and development: a systematic literature review." European Journal of Training and Development 47.7/8 (2023): 677-694.

[9] Alqurashi, Daher Raddad, et al. "Exploring the impact of artificial intelligence in personalized content marketing: a contemporary digital marketing." Migration Letters 20.S8 (2023): 548-560.

[10] Ruiz-Rojas, Lena Ivannova, et al. "Empowering education with generative artificial intelligence tools: Approach with an instructional design matrix." Sustainability 15.15 (2023): 11524.

[11] Tejani, Ali S., et al. "Artificial intelligence and radiology education." Radiology: Artificial Intelligence 5.1 (2022): e220084.

[12] Zaman, Khansa. "Transformation of marketing decisions through artificial intelligence and digital marketing." Journal of Marketing Strategies 4.2 (2022): 353-364.

[13] Lu, Zhiqin, and Inyong Nam. "Research on the influence of new media technology on internet short video content production under artificial intelligence background." Complexity 2021.1 (2021): 8875700.

[14] Lancaster, Thomas. "Artificial intelligence, text generation tools and ChatGPT–does digital watermarking offer a solution?." International Journal for Educational Integrity 19.1 (2023): 10.