ORIGINAL RESEARCH

Service quality in football tourism: an evaluation model based on online reviews and data envelopment analysis with linguistic distribution assessments

Adjei Peter Darko[1] · **Decui Liang**[1] · **Yinrunjie Zhang**[1] · **Agbodah Kobina**[2]

Abstract

The emergence of sports tourism has compelled sports managers to rethink the management and improvement of sports facilities. Through service quality analysis, sports managers can identify the strengths and weaknesses of their activities for possible advancement. Hence, this study aims to develop a decision support model based on integrating online reviews and data envelopment analysis to measure the degree of tourist satisfaction and provide benchmarking goals for service improvement. The proposed model employs text mining techniques to discover service quality attributes from text reviews. According to the discovered service quality attributes, we conduct sentiment analysis to reveal the sentiment polarities of the text reviews. Then, we refine the polarities and ratings of online reviews into linguistic distribution assessments. Furthermore, we develop a linguistic distribution output-oriented non-discretionary bestpoint slack-based measure (BP-SBM) to compute the degree of tourist satisfaction and benchmarking goals. The linguistic distribution output-oriented non-discretionary BP-SBM can handle both positive and negative data values, thus overcoming the flaws of the traditional model. Meanwhile, the proposed decision support model investigates how the service-quality attributes interact to provide improvement pathways for an underperforming stadium based on association rule mining. We test the applicability of the proposed decision support model on some Elite stadia in Europe.

Keywords Online reviews · Sports stadium service quality · Benchmarking analysis · Quality association · Linguistic distribution assessments

---

✉ Decui Liang
decuiliang@126.com

Adjei Peter Darko
paadarko@yahoo.com

Yinrunjie Zhang
yinrunjie-zhang@outlook.com

Agbodah Kobina
kobinaagbodah@yahoo.com

1 School of Management and Economics, University of Electronic Science and Technology of China, Chengdu 610054, China

2 Department of Applied Mathematics, Koforidua Technical University, Koforidua, Ghana

1 Introduction

Sports teams have recognized the need to open up their facilities to the general public through tourism to gain extra revenue and positive public relations. Sports and tourism are the two fastest expanding tourism industries (Kurtzman & Zauhar, 2003). Every aspect of sports, travel and tourism is now linked (Takata & Hallmann, 2021). Sports tourism is beneficial to tourism growth (Cho et al., 2019). Individuals, regions, and countries all benefit from sports tourism. Tourists may relate to their team's history via sports tourism. Others, on the other hand, consider sports tourism as the continuation of a national, regional, or personal tradition and identity. Sports tourism is becoming an essential component of certain local development programs, particularly in interior areas and territories where it may be combined with cultural tourism and the natural environment, for example, to produce fascinating outcomes. From 2021 to 2030, the worldwide sports tourism market is expected to grow at a CAGR of 16.1% from $323,420.0 million in 2020 to $1,803,704.0 million (Aniket & Roshan, 2021). This illustrates that sports tourism's economic and social effect extends well beyond the actual athletic event. Locals and tourists alike benefit from the social and economic benefits it creates.

Conventional wisdom held that sports tourism was limited to spectator vacations. Redmond made this remark between 1990 and 1991, saying that spectators are only one kind of sports tourist (Redmond, 1981, 1990). Redmond categorized sports fans as "actors" of sports tourism since they are the ones who visit sports venues like historic arenas and museums dedicated to sports legends. Gibson (1998) definition of sports tourism is regarded as the most thorough and accurate. He defined it as "Leisure-based travel that takes individuals temporarily outside of their home communities to participate in physical activities, to watch physical activities, or to venerate attractions associated with physical activities" (Gibson, 1998). Based on this definition, three types of sports tourism are identified: (1) active sports tourism, (2) event sports tourism, and (3) nostalgia sports tourism. Over the last several years, there has been a steady rise in studies related to sports tourism. As a result, most studies have focused on today's big sports shows (Niu & Zhang, 2021). It's becoming more common to see sporting facilities as tourist attractions in their own right, though (Vegara-Ferri et al., 2020). Hence, in this research, we delve into nostalgia sports tourism. A well-known definition of nostalgia sports tourism includes visiting sports stadia/arenas, halls of fame, and other locations related to a sport's history (Gibson, 1998). Particularly, this research focuses on football stadium tours, a guided tour behind the stadium scenes.

A growing number of sports and tourism academics are attempting to understand better elements that encourage people to participate in football stadium tours (Takata & Hallmann, 2021). Like in many other types of sports tourism, the desire to return is crucial, particularly for tourism focused on sports heritage (Chalip & McGuirty, 2004). Many studies have shown that repeat visitors stick around for a more extended period and are more likely to share favorable word-of-mouth about the site (Oppermann, 2000). In addition to representing visitors' loyalty, the desire to return in the future might assist in forecasting their future behavior (Chen & Rahman, 2018). Because many sporting events and tourism places compete for the same sports tourists, it is critical to know and promote visitors' intentions to return. The quality of service is among the numerous factors scholars have studied concerning customer desire to return (de Matta & Lowe, 2021). Managers who want to restore the image of their sporting facilities via efficient marketing techniques must focus on service quality. There is a correlation between greater satisfaction levels and better economic returns (de Matta & Lowe, 2021). Quality service encourages sports fans to travel to cheer on their favorite teams, visit sports museums

and halls, and form bonds with other fans who share their goals (Fairley, 2003). Tourists expect a higher service quality from tourism managers since going for such experiences necessitates more time and money than participating in other sports-related activities (Cho et al., 2019).

Tourists often use tourism websites and social media to select tourist attractions based on other visitors' experiences. Text reviews and ratings on travel websites and social media give a wealth of data that may be used to measure the quality of services offered (Yang et al., 2021). In the case of football stadium tours, many tourists write evaluations on Tripadvisor.com to share their thoughts and emotions. Managers may gauge tourist satisfaction and make adjustments to their services due to the feedback they get from customers. Identifying service quality-related factors is critical for increasing tourist satisfaction. Because of this, this research uses online reviews to model service quality and assess football stadium attractions. This enables service managers to identify areas for improvement in service quality immediately.

Evaluating the service quality of football stadium tours based on online reviews involves the consideration of several conflicting inputs and outputs; hence can be achieved through utilizing data envelopment analysis (DEA). Data envelopment analysis (Banker et al., 1984) is a non-parametric modeling method that utilizes few assumptions to solve performance-related problems with intricate relationships between several inputs and outputs. DEA has been known to be a robust optimization technique to handle service quality measurement (Park & Lee, 2021). Recently, Park and Lee (2021) integrated online reviews and DEA to benchmark hotel service. The authors developed a decision-support framework for hotel managers to comprehensively estimate the degree of guest satisfaction (i.e., service-quality measure) together with benchmarking guidelines on service quality improvement. As a result, combining online reviews with DEA is a step on the right path. Nonetheless, online reviews are riddled with inconsistencies and imprecision. Consequently, their usage in DEA may lead to erroneous results. Therefore, there is a pressing need to model online review uncertainties and combine them with DEA for better benchmarking analysis.

To overcome the issue above, it is essential to address the issue of characterizing online football stadium reviews. Herrera et al. (1995) replaced linguistic variables with a linguistic term set (LTS) to explain imprecise knowledge. However, employing LTS to manage them is problematic because of many online comments. However, Zhang et al. (2014) suggested linguistic distribution assessments (LDAs) that may be utilized to define huge linguistic variables. LDAs are strong sets for statistically summarizing group linguistic assessments. LDAs have been used to solve various decision-making difficulties because of their high feasibility in displaying qualitative and quantitative data (Nie et al., 2020; Wu et al., 2021a, 2021b; Wu et al., 2021a, 2021b; Xiao et al., 2020). LDAs are more versatile and resilient than current linguistic representations since they can convey individual and group linguistic judgments. As a result, we employ LDAs to define online football stadium reviews in this research. To our knowledge, LDAs are yet to be used in DEA analysis to represent online reviews.

This study aims to propose a data-driven decision support framework proficient in diagnosing football stadium online reviews for benchmarking and providing improvement pathways for service quality. First, it utilizes natural language processing techniques to extract the key service quality attributes from online reviews. Afterward, we perform sentiment analysis to reveal the polarity of tourists' opinions towards different service quality attributes. In this instance, we apply LDAs to refine the online reviews of the tourists. Then, we develop a linguistic distribution DEA model to measure tourist satisfaction and perform benchmarking analysis for quality improvement. The LDAs can model tourists' opinions from positive, neutral, and negative perspectives. Hence, the traditional DEA model cannot handle this kind

of data. As a result, this study employs a non-positive slack-based measure (SBM) of DEA to overcome this data challenge. In addition, the study employs association rule mining to design a service quality improvement pathway.

Against this backdrop, the study contributes to the literature in the following ways:

1. Acquisition of service quality attributes. Attributes are used as rules to evaluate the alternatives. Hence, the selection of attributes is considered essential. Based on online text reviews, we utilize robust natural language processing (NLP) algorithms for mining key service attributes, which serve as data for the DEA analysis instead of depending on traditional service quality indices.
2. Description of tourists' opinion. Online textual reviews are enormous in number and disordered in structure. Imprecision of tourists' opinions leads to incomplete data, affecting the decision results. Hence, how to describe tourists' opinions to be involved in decision analysis is of vital importance. This study defines some new rules to describe online reviews using LDAs. Then, we design a novel score function of LDAs.
3. Development of measurement model. Service quality measurement is essential for ensuring the revisit of tourists and the attraction of new tourists. Therefore, developing a robust technique for benchmarking tourist satisfaction is crucial. This study proposes a linguistic distribution output-oriented non-discretionary bestpoint slack-based measure (BP-SBM) to model and benchmark the service quality of football stadia. The proposed method can model the ignorance and incompleteness of online reviews that other DEA models ignore. Meanwhile, the study exploits the association rule mining to design an improvement pathway for managing sports tourism service quality.

The rest of the paper is summarized: Sect. 2 reviews the existing literature on service quality, DEA, and LDAs. Section 3 discusses the variables for the DEA, data preparation designs a novel score function for LDAs and proposes an output-oriented non-discretionary linguistic distribution SBM of DEA for benchmarking analysis. Also, we conduct quality associations based on the association rule mining approach. In Sect. 4, we output and discuss the results of the tourist satisfaction measure, the benchmarking analysis, and the quality association. Also, we conduct a comparative analysis in this section. Section 5 throws light on some implications of the study. Section 6 presents the conclusion of the paper and elaborates on future studies.

2 Related works

2.1 Service quality in sports facilities

In the tourism sector, most experiences are created for tourists (Mariani & Visani, 2019). As a result, facilities and services are constantly evolving (Lau et al., 2021). Tourists' evaluations and satisfaction with such services, in particular, are influenced by service-based encounters (Edensor et al., 2021). Tourists' subjective reactions to exchanges with services and service providers, both direct and indirect, are altered due to these encounters (Edensor et al., 2021). Tourists must have physical comforts in addition to the services they get. Social media, blogs, wikis, and other collaborative media have greatly simplified the examination of tourist actions and impressions. Online reviews are a popular tool for tourists to share their thoughts on tourist attractions they have visited. They disseminate a large quantity of information based on a wealth of data and encourage the production of new knowledge (Darko & Liang,

2022b). Using online reviews to enhance service quality, marketing, and the environment is a fantastic potential for tourist businesses (Guo et al., 2017).

User satisfaction may be significantly influenced by how sports facilities are administered (Cho et al., 2019). This field's relevant research may be divided into functional and service management streams. Studies focusing on the functional attributes of sports facilities utilize technical approaches to evaluate specific building services management characteristics such as air quality (Bunds et al., 2019; Silva & Ricketts, 2016), thermal condition (Bonser et al., 2020; Losi et al., 2021), and heating energy consumption (Bonser et al., 2020; Sofotasiou et al., 2014). Sports facilities' technical performance was evaluated using engineering investigations (such as computer simulations and physical measurements) in these studies, often undertaken from a financial or environmental aspect. The other studies focused on the service aspect of sports facilities use user-centric analytic approaches to evaluate the management of sports facilities, such as benchmarking analysis (Iversen, 2015; Liu et al., 2009), management efficiency (Li & Li, 2021; Ramchandani et al., 2018), and quality management (Ramchandani & Taylor, 2011).

In the administration of sports facilities, service quality is critical. Several studies have investigated how service quality affects tourist satisfaction and return propensity. For example, in Mazandaran province, Ghasemi and Sanaei, (2015) investigated the link between attractive sports facilities and customer satisfaction as mediated by service quality. Álvarez-García et al. (2019) investigated the impact of the characteristics that allow for the grading of service quality perceived by sports and health facilities users on their satisfaction with the service received. Ibrahim et al. (2020) investigated the effect of service quality dimensions in sports facilities on use intention.

Most research on service quality in sports facilities has employed a quantitative approach, including surveys and scale development methodologies. Survey-based service quality methods have certain disadvantages, such as being time-consuming, requiring a significant financial commitment, and having lower accuracy owing to a lack of continuous period assessment (Darko & Liang, 2022a). In conclusion, survey-based service quality instruments are ineffective for rapid and proactive decision-making. Hence, the introduction of a novel methodology to overcome the drawbacks of the survey-based methods is very crucial (Weed & Bull, 2009). Processing online reviews has been identified as a potentially strong alternative to assessing survey data (Park & Lee, 2021). As a result, there is a need for a paradigm shift toward analyzing service quality in sports facility management by employing tourist experiences through online reviews.

Indeed, little research has been conducted to explore tourist experiences in sports tourism, notably football stadium tourism. For example, Edensor et al. (2021) evaluated tourist experiences inside football stadia in the 2019–2020 season using TripAdvisor reviews from 44 English football stadia. The authors analyzed online reviews to enhance the visitor experience in tourist sites, examining multi-dimensional encounters and establishing various dimensions that assist management in comprehending the complexity of their offers. The research backed earlier results regarding how stadiums act as pilgrimage and historical sites. In addition, two new visitor experience characteristics were unveiled, demonstrating how stadiums serve as restaurants and have practical and functional features that are essential to the tourist experience. With insights from the study of Edensor et al. (2021), we extend the application of online football stadium reviews to investigate tourist satisfaction with football stadium service quality.

2.2 DEA studies

For many years, one of the main goals of DEA has been to assess the relative efficiency of service-producing units (Zhu, 2022). As a result, several studies have used it to evaluate customer satisfaction and service quality. DEA is a non-parametric multiple criteria decision-making technique that comprises the identification of decision-making units (DMUs), a set of inputs and outputs for each DMU and the measurement of efficiency values of DMUs.

DEA has been used in some tourism-related research for benchmarking purposes. For example, Sellers-Rubio and Casado-Díaz (2018) used a two-stage double bootstrap data envelopment analysis approach to evaluate the performance of the Spanish hotel business. According to the findings, there is a substantial correlation between the environmental factors studied and the overall inefficiency of hotels in Spain. Yin et al. (2020) provided a two-stage hotel network structure model based on resource-based perspective theory (i.e., operations stage and marketing stage). A bi-objective model built on DEA was proposed to assess the hotel's efficiency from two interdependent phases. Several efficiency studies use historical data as the input and output variables of the DEA model, similar to the works mentioned above.

Other research has eschewed historical data favoring survey data when developing DEA models. For instance, Lee and Kim (2014) proposed a pure output-oriented Banker–Charnes–Cooper (BCC) approach for quantifying and benchmarking service quality across five dimensions (i.e., tangibles, reliability, responsiveness, assurance, and empathy) using established service performance criteria. Najafi et al. (2015) used the perceived service quality index (PSQI) as a single measure to assess the multiple-item service quality concept. The PSQI is calculated using a slack-based measure (SBM) of efficiency with constant inputs.

Recent advancements in social media and internet websites have resulted in massive amounts of open-source data for decision-making. To better understand tourist experiences, several studies used internet reviews as a data source for DEA models, namely ratings and text reviews. Mariani and Visani (2019) published ground-breaking research that explored the importance of online customer satisfaction rating scores for calculating efficiency scores and enhancing the discriminative power of DMUs (hotels) on hotel performance. They sought to close the gap between online rating scores and hotel performance by applying a classical input-oriented BCC-DEA model to a sample of various Italian hotel categories. The model included three input variables (rooms, employees, and net operating expenses) and two output variables (revenue and online rating scores). Park and Lee (2021) created a decision-support framework based on DEA and sentiment opinions to thoroughly evaluate guest satisfaction (i.e., service-quality metrics and benchmarking suggestions for improvement).

It is evident that different data sources are used in the DEA model for benchmarking analysis. However, online reviews better describe tourists' experiences as they provide rich information. Hence, integrating online reviews with DEA is in the right direction. The pioneering works (Mariani & Visani, 2019; Park & Lee, 2021) provide the basis for using online reviews in the DEA model. Nevertheless, online reviews are usually characterized by ambiguous and incomplete information; hence their use in decision-making may result in poor outcomes. Therefore, we aim to bridge this gap by modeling online reviews' uncertainties and integrating them with DEA for better benchmarking analysis.

2.3 LDAs

With the rising ambiguity of information and decision-makers' subjective cognition in recent years, it has become more challenging to employ exact values for alternative assessments throughout the decision-making process. As a result, decision-makers are more likely to use language phrases to describe their intuitive views regarding assessment difficulties. Numerous linguistic evaluations in real-world decision-making, such as online reviews on the tourist website (TripAdvisor.com), take the form of distributed assessments (Yu et al., 2018). As a result, linguistic distribution assessments (LDAs) are seen as an appropriate representation of the subjective ambiguity and objective distributed assessment of decision makers' viewpoints. Compared to a variety of other linguistic models, LDAs can effectively represent uncertain information while maintaining the integrity of group linguistic information. To more accurately and completely express group views, it is critical to use LDAs to convey group linguistic assessments.

The LDAs-based expressions have shown productive results due to their great practicality in representing qualitative and quantitative information. For example, through linguistics, Xiao et al. (2020) investigated how to handle individual semantics and consensus in large-scale group decision-making. Nie et al. (2020) used LDAs to construct a cloud-based quality function deployment (QFD) model in conjunction with the TODIM approach to improve the quality of healthcare services. Again, Wu et al. (2021a, 2021b) created a quantum framework for simulating interference effects in multiple criterion group decision-making for linguistic distributions. Wu et al., (2021a, 2021b) suggested a model of multi-criteria group decision-making based on linguistic distributions that integrate extended TODIM with quantum decision theory.

A detailed examination of the available literature demonstrates that the LDAs model is employed to reflect individual rather than group linguistic judgments. However, many LDAs are group evaluations, such as movie ratings, product praise ratings, and emotional views. Only a few studies have examined group linguistic assessments via online reviews. For instance, Yu et al. (2018) used LDAs to develop a multi-criteria decision-making model for hotel selection. The authors used their suggested model to convert online reviews into LDAs to assess hotels and choose the ideal one. According to Yu et al. (2018), group linguistic assessments such as online reviews may be represented as LDAs. Because the LDAs' representation of linguistic information is more convenient and thorough, this research uses the LDAs to express the distributed group linguistic evaluation information, which is more consistent with the current environment.

In what follows, we present some fundamentals of the LDAs based on the results of Zhang et al. (2014) and Yu et al. (2018), which can serve as the theoretical background for the rest of the paper.

Definition 1. Let $S = \{s_\delta | \delta = -\tau, \ldots, -1, 0, 1, \ldots, \tau\}$ be a linguistic term set (LTS); then an LDA is defined as follows:

$$Ld = \left\{ \left(s^{(\tau)}, \beta^{(\tau)} \right) | s^{(\tau)} \in S, \beta^{(\tau)} > 0, \tau = 1, 2, \ldots, |Ld|, \sum_{\tau}^{|Ld|} \beta^{(\tau)} = 1 \right\}, \quad (1)$$

where $s^{(\tau)}, \beta^{(\tau)}$ is the τ − th element in the LDA and comprises a linguistic term $s^{(\tau)}$ and its symbolic proportion $\beta^{(\tau)}$. The term $|Ld|$ represents the number of elements in the LDA.

Let $S = \{s_\delta | \delta = -\tau, \ldots, -1, 0, 1, \ldots, \tau\}$ be an LTS and $Ld = \left\{(s^{(\tau)}, \beta^{(\tau)}) | s^{(\tau)} \in S, \beta^{(\tau)} \geq 0, \tau = 1, 2, \ldots, |Ld|, \sum_{\tau}^{|Ld|} \beta^{(\tau)} = 1\right\}$ be an LDA, then a score function value Ld is given as $E(Ld) = \sum_{\tau=1}^{|Ld|} \beta^{(\tau)} \Phi(s^{(\tau)})$, where the term $\Phi(s^{(\tau)})$ is the subscript of the linguistic term $s^{(\tau)}$.

For any two LDAs Ld_1 and Ld_2, if $E(Ld_1) > E(Ld_2)$, then Ld_1 is greater than Ld_2; if $E(Ld_1) = E(Ld_2)$, then Ld_1 and Ld_2 have the same score; if $E(Ld_1) < E(Ld_2)$, then Ld_1 is lesser than Ld_2.

3 A data-driven decision support framework

This study proposes a comprehensive data-driven decision support framework to adeptly diagnose football stadium online reviews for benchmarking and improving service quality management pathways. The proposed framework comprises four phases: service quality attributes extraction, sentiment analysis and transformation, benchmarking analysis, and service quality improvement analysis. Figure 1 summarizes the comprehensive decision support model.

3.1 Variable selection

According to the suggested linguistic distribution SBM of DEA, we need to define the input and the output components to calculate the degree of tourist satisfaction. Two dimensions, firm reputation and service quality, are significant drivers in gauging customer happiness. Firm reputation is defined as a stakeholder's total judgment of a firm over time which may be based on personal experiences or the experiences of third parties. TripAdvisor's 2018 research estimated that 97% of company owners regard online reputation management as crucial to their companies (Erskine, 2018). Other firm characteristics may be used to measure reputation, such as tourists per year, annual revenue, or the firm's age. However, this study focuses on using tourist experiences through online reviews other than historical data to measure online reputation. A firm's online reputation is reflected in its ratings (Luca & Reshef, 2021). Online rating is one of the crucial pieces of information in a firm's reputation since it provides experiences from prior customers. Hence, some research has used online ratings as a proxy for firms' reputations (Chiles, 2021; Luca & Reshef, 2021). For instance, Luca and Reshef (2021) evaluated the influence of pricing on business reputation using internet ratings to measure firm reputation. Also, Chiles (2021) studied the influence of required extra fees on business reputation in the U.S. hotel market using data gathered from two major online travel sites. The author used disparities in surcharge disclosure across booking channels to find concealed "resort fees" causal influence on passenger evaluations. Based on the conversations thus far, we use online ratings as the proxy for the stadium's reputation in the suggested approach.

On the other side, the quality of service is a significant indicator of tourist satisfaction (Masriki & Frinaldi, 2021; Santos et al., 2020). Numerous researchers have demonstrated a link between service quality and tourist satisfaction (e.g., Masriki & Frinaldi, 2021; Moro et al., 2020; Santos et al., 2020; Su & Teng, 2018). In recent years, online textual reviews have developed into a valuable data source for assessing service quality. Textual reviews are used
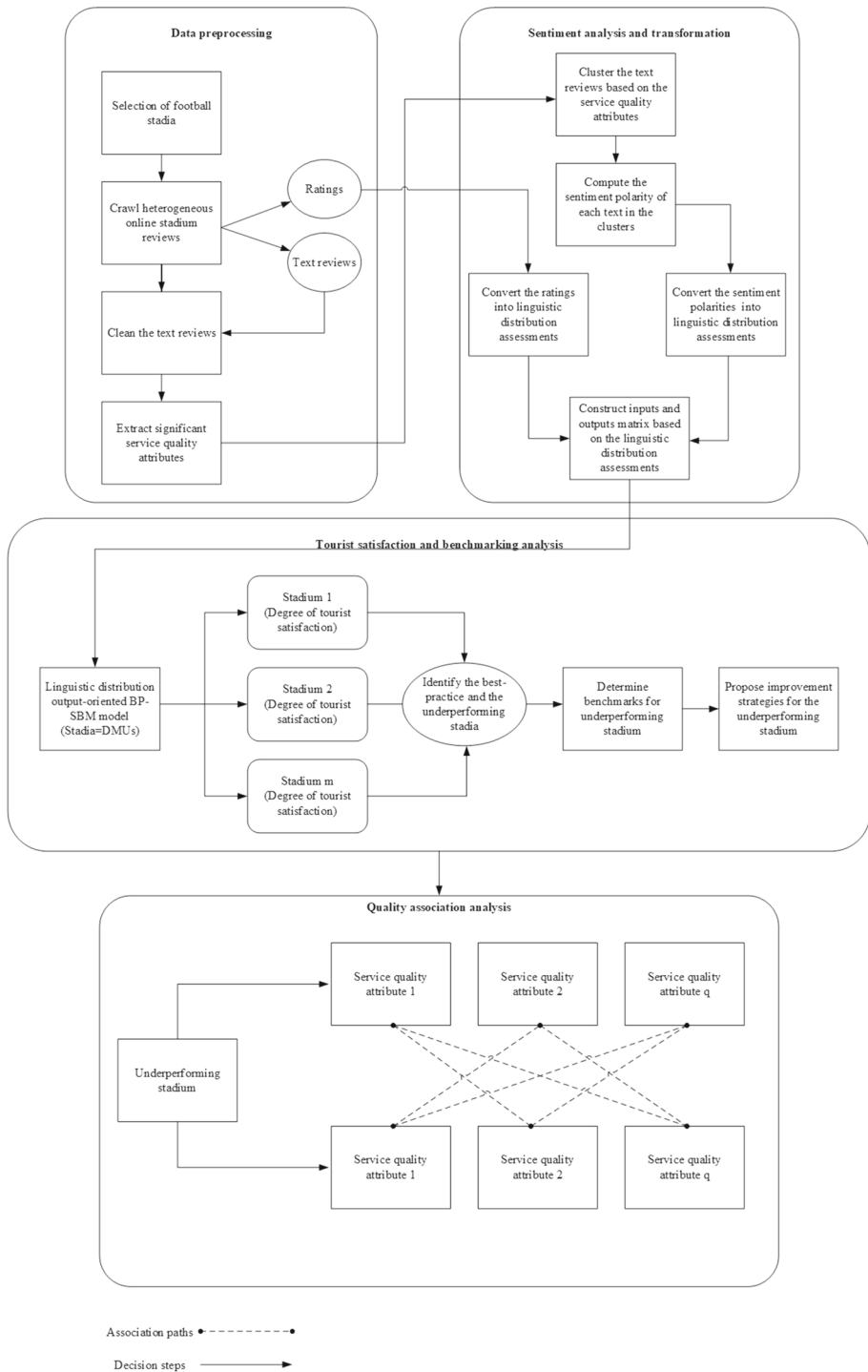
Fig. 1 Framework of the proposed data-driven decision support model

to identify service quality characteristics, and sentiment analysis is used to determine overall imprecision (Nilashi et al., 2021a, 2021b). For instance, Nilashi et al. (2021a, 2021b) studied travelers' satisfaction in Malaysian hotels during the COVID-19 outbreak through online customers' reviews. Moreover, the authors examined whether service quality during COVID-19 has an impact on hotel performance criteria and consequently customers' satisfaction. As an aspect of smart tourism Kim et al. (2017) applied the sentiment analysis method to analyze the quality of tourism destinations. Also, Thu et al. (2021) presented a method to measure guest satisfaction based on sentiment lexicon that is developed for hospitality domain. Based on these previous studies, it is possible to analyze service quality of football tourism using online reviews and sentiment analysis. Therefore, using the procedures mentioned in Sect. 3.2, we extract essential service quality attributes from stadium online text reviews. We analyze the service quality attributes to ascertain their sentiment polarity for further analyses.

There is substantial proof that a firm reputation significantly impacts consumers' evaluative opinions, such as quality assessments (Andreassen & Lindestad, 1998; Darden & Schwinghammer, 1985). A firm reputation is essential to service quality (Anderson & Sullivan, 1993; Blanchard & Galloway, 1994). Gronroos (1988), for example, presented three determinants of service quality: technical excellence, professionalism and abilities, and image. The image dimension is concerned with one's reputation and credibility. Furthermore, Bloemer et al. (1998) researched image-related problems in banks and discovered that a bank's favorable brand image considerably affects perceived service quality. In other words, brand image (reputation) is an essential indicator of service quality. Wu (2011) investigated the influence of hospital brand image on service quality, patient satisfaction, and loyalty. The findings revealed that hospital brand image improves service quality, patient satisfaction, and loyalty. In addition, Vimla and Taneja (2021) provided a conceptual framework in which they found that hospital brand image positively impacted service quality.

As a result, the research hypothesizes that stadium reputation (input) influences the quality of service (output) offered by stadium management, based on a link between company reputation and service quality that the literature has experimentally supported. As a result, in the proposed DEA model, we use online ratings to proxy stadium reputation and sentiment polarity as a surrogate for service quality qualities. In online decision-making, we propose that ratings and feelings have distinct proximities. Apart from the fact that numeric ratings may not wholly capture the "polarity of information in the text evaluations" (Hu et al., 2014), we believe that feelings may play a different function in the decision-making process than ratings. Tourists may use various types of information throughout different stages of the selection process since search, assessment, and choosing in an online environment may be highly complex. Hu et al. (2014) investigated the function of review sentiment in moderating the impact of product ratings on sales. The authors discovered that ratings did not affect sales; however, average sentiments negatively and substantially affect sales. Their findings show that ratings indirectly influence sales rank through sentiments, whereas sentiments directly impact sales rank. Based on these statements, we conclude that online ratings and sentiments may be included in the proposed DEA model since they play distinct roles in the decision process.

3.2 Data preparation

Data preparation involves two main stages: extracting online football stadium reviews and data preprocessing. Natural language processing (NLP) techniques are employed to accomplish the data preparation phase. A detailed explanation of each step is discussed below.

3.2.1 Crawling the online reviews

This step crawls online football stadium reviews from Tripadvisor.com (https://www.tripadvisor.com), a popular travel website globally. We scrape online football stadium reviews using python programming based on selenium. Typically, OCRs contain much information such as reviewer name, review date, text reviews, numerical reviews (ratings), etc. Therefore, the required information can be obtained by utilizing a python crawler.

3.2.2 Preprocessing the textual reviews

First, we leverage RegEx (Malik et al., 2021) and other normalization techniques to clean the text reviews by eliminating special characters (e.g., &, %, @, *, #, etc.) and unclarified numbers. SymSpell (Garbe, 2019) is used to correct typo errors, and the text reviews are translated into lower cases. Next, we perform tokenization based on SpaCy (Goyal et al., 2018) to break down the text reviews into words. Afterward, we delete English stop words (e.g., "I," "am," "what," "is," etc.) from the tokenized words. Then, to bring single word concepts together, stemming techniques are used, in which multiple versions of the word, such as "absolutely," "absolute," and "absoluteness," are reduced into the single token "absolute."

3.3 Extraction of service quality attributes

We utilize the Latent Dirichlet allocation (Blei et al., 2003) topic modeling approach to grasp the hidden attributes of service quality from the textual reviews. In other words, we discover per-attribute word distributions (i.e., per-topic word distributions) for identifying service quality attributes and per-review attribute distributions (i.e., per-document topic distributions) for vectorizing each review for further analyses. Note that topics extracted from the Latent Dirichlet allocation model correspond to service quality attributes. Based on the Gibbs Sampling formula, all the words in the textual reviews are iterated, reorganized, and each word's feature distributions are updated. The outcome of the trained Latent Dirichlet allocation model includes a "Topic-word" matrix. We integrate similar topics and finalize the topic's name, representing significant service quality attributes $F_r (r = 1, 2, \ldots, q)$

Let $T_i (i = 1, 2, \cdots, v)$ denote a textual review and $F_r (r = 1, 2, \ldots, q)$ represent an attribute. Then, we construct the text-attribute matrix based on the outcome of the Latent Dirichlet allocation model. Each value in the text-attribute matrix represents how a textual review belongs to a specific attribute. For each textual review, we can select the max value from its text-attribute probability vector as the primary attribute for the text. Hence, each service attribute is a cluster of several textual reviews.

3.4 Translation of sentiment scores into LDAs

In this subsection, we conduct sentiment analysis on the cluster of service quality attributes containing a group of similar textual reviews. Based on these textual reviews, we use the VADER (Hutto & Gilbert, 2014) sentiment analyzer to detect the sentiment score of each textual review. Let φ represent the sentiment score where $\varphi \in [-1, 1]$. In this case, 1 represents the most positive opinion, while -1 denotes the most negative opinion. Based on sentiment scores, stadium managers can interpret tourist experiences regarding each key service quality attribute. The sentiment scores have three levels, i.e., positive, neutral, and negative.

Subsequently, we employ the LDAs to help summarize the sentiment scores regarding the service quality attributes. The LDAs help quantify the accurate reflection of tourist feelings and overcome the uncertainty in benchmarking tourist satisfaction. According to the framework of the LDAs, we describe the sentiment scores using five linguistic labels:

(a) If φ is between -1 and -0.5, it can be denoted as very negative (s_{-2});
(b) If φ is between -0.5 and 0, it can be marked as negative (s_{-1});
(c) If φ is equal to 0, it can be denoted as general (s_0);
(d) If φ is between 0 and 0.5, it can be designated as positive (s_1);
(e) If φ is between 0.5 and 1, it can be represented as very positive (s_2).

On the other hand, we translate the rating score provided by the tourist as well into five linguistic labels:

(a) If μ is equal to 1, it can be denoted as very negative (s_{-2});
(b) If μ is equal to 2, it can be characterized as negative (s_{-1});
(c) If μ is equal to 3, it can be denoted as general (s_0);
(d) If μ is equal to 4, it can be marked as positive (s_1);
(e) If μ is equal to 5, it can be represented as very positive (s_2).

Meanwhile, we calculate the frequency of each linguistic term and obtain the LDAs.

The LDAs obtained based on the sentiment scores are used to evaluate the service quality attributes. Similarly, the LDAs generated from the rating scores are utilized to measure a firm reputation. Subsequently, we introduce the LDAs into the DEA framework and propose a benchmarking methodology for analyzing football tourism service quality.

3.5 Linguistic distribution SBM of DEA

In this part, we develop an approach for benchmarking sports stadia. The benchmarking study evaluates each stadium's degree of tourist satisfaction and offers a target stadium for other stadia with low tourist satisfaction (Park & Lee, 2021). SBM of DEA (Tone et al., 2020) is an appropriate approach for performing the benchmarking study since it directly models the input or output slacks without assuming proportional changes of inputs or outputs. Hence, we integrate it with LDAs and propose an improved linguistic distribution SBM of DEA.

This research assumes that LDAs have a different impact on tourist satisfaction and benchmarking strategies. Specifically, the LDAs over the multiple outputs and inputs for the SBM DEA may be positive, negative, or zero. Therefore, we apply the Basepoint (BP)-SBM DEA (Tone et al., 2020) model to handle this problem. The BP-SBM DEA assumes that all inputs or outputs are discretionary, i.e., controlled by each DMU manager, and can be changed at their will. However, in any realistic case, non-discretionary inputs or outputs may exist beyond managers' control. For example, location can be considered non-discretionary. This is because a location's selection for a particular project, such as buildings, involves several decision factors such as accessibility, security, competition, and growth potential. Once established, a firm often chooses to remain in its initial location owing to external economies of scale. A place that has been connected with a particular sector acquires specialized skills and knowledge over time. A relocation choice entails potentially large expenditures, such as equipment transfer. Hence, selecting a location is a long-term decision and may not be easily altered by managers. These non-discretionary inputs or outputs are usually excluded from the benchmarking analysis but aid in measuring efficiency.

Estimating the LDAs, and the quest to measure the ideal service quality (i.e., the extent of tourist satisfaction) over multiple key service attributes require developing an output-oriented BP-SBM DEA with multiple outputs and less input. In this paper, the LDAs of service attributes are used as outputs, while the LDAs of rating used as a proxy for stadium reputation are considered an input. Tone et al. (2020) introduced the following BP-SBM DEA as Model 1 to handle non-positive exact numbers.

Model 1

$$\theta_o = \min \frac{1 - \frac{1}{n} \sum\limits_{i=1}^{n} \frac{\eta_{io}^-}{\bar{x}_{io}}}{1 + \frac{1}{h} \sum\limits_{i=1}^{h} \frac{\eta_{io}^+}{\bar{y}_{io}}}$$

$$s.t. \begin{cases} \sum\limits_{j=1}^{m} \lambda_j \bar{x}_{ij} + \eta_{io}^- = \bar{x}_{io}, & (i = 1, 2, \ldots, n) \\ \sum\limits_{j=1}^{m} \lambda_j \bar{y}_{ij} - \eta_{io}^+ = \bar{y}_{io}, & (i = 1, 2, \ldots, h) \\ \eta_{io}^- \geq 0, \eta_{io}^+ \geq 0, \lambda_j \geq 0 \ (j = 1, 2, \ldots, m) \end{cases}$$

where $\bar{x}_{ij} = x_{ij} - x_i^{\min} > 0 (\forall i, j)$ and $\bar{y}_{ij} = y_{ij} - y_i^{\min} > 0 (\forall i, j)$ are the modified base point used to translate the negative inputs and outputs. The optimal solution of Model 1 is $\left(\theta_o, \lambda_j^*, \eta_{io}^{*-}, \eta_{io}^{*+}\right)$.

Inspired by the above idea, we modify the BP-SBM DEA model in the linguistic distribution environment by introducing LDAs into the BP-SBM DEA. The fundamental basis of this research lies in how stadium managers can improve tourist satisfaction through service quality attributes. Hence, we adopt the output-oriented variant of the BP-SBM DEA to measure the degree of tourist satisfaction. Again, we argue that some service quality attributes are beyond the control of stadium managers. Therefore, we introduce non-discretionary outputs in the BP-SBM DEA.

Definition 2. For a set of inputs $X = \{x_i | i = 1, 2, \ldots, n\}$ and a set of outputs $Y = \{y_r | r = 1, 2, \ldots, q\}$ of the stadium k ($k = 1, 2, \ldots, m$), suppose that the values of input are contained in a linguistic distribution input matrix $(LDIM)$ $I = (\Psi_{ik})_{n \times m}$, and the outputs are contained in a linguistic distribution output matrix $(LDOM)$ $O = (\Psi_{rk})_{n \times q}$.

According to the proposed output-oriented non-discretionary BP-SBM DEA, when the input–output data contain both LDAs, for the assessed DMUs, the output-oriented non-discretionary BP-SBM DEA in a linguistic distribution environment is defined as Model 2, i.e.,

Model 2

$$\theta_k = \min \frac{1 - \frac{\varepsilon}{n} \sum\limits_{i=1}^{n} \frac{\eta_{ik}^-}{\Psi_{ik}}}{1 + \frac{1}{h} \sum\limits_{r=1}^{h} \frac{\eta_{r^D k}^+}{\Psi_{r^D k}}}$$

$$s.t. \begin{cases} \sum\limits_{j=1}^{m} \lambda_j \overline{\Psi}_{ij} + \eta_{ik}^- = \overline{\Psi}_{ik} & (i \in I) \\ \sum\limits_{j=1}^{m} \lambda_j \overline{\Psi}_{r^D j} + \eta_{r^D k}^- = \overline{\Psi}_{r^D k} & (r^D \in O) \\ \sum\limits_{j=1}^{m} \lambda_j \overline{\Psi}_{r^{ND} j} = \overline{\Psi}_{r^{ND} k} & (r^{ND} \in O) \\ \eta_{ik}^- \geq 0, \eta_{r^D k}^+ \geq 0, \lambda_j \geq 0 & (j = 1, 2, \dots, m) \end{cases}$$

In Model 2, ε is a negligible positive integer; r^D and r^{ND} represent discretionary and non-discretionary outputs in the $LDOM$, η_{ik}^- and $\eta_{r^D k}^+$ are the slack variables connected with the inputs and the discretionary outputs, respectively; $\overline{\Psi}_{ij} = \Psi_{ij} - (\Psi_i)^{\min} > 0 (\forall i, j)$; $\overline{\Psi}_{rj} = \Psi_{rj} - (\Psi_r)^{\min} > 0 (\forall r, j)$; λ_j is the weight assigned to the jth stadium. Thus, the optimal solution of Model 1 is $\left(\theta_k, \lambda_j^*, \eta_{ik}^{*-}, \eta_{r^D k}^{*+} \right)$.

To solve Model 2, we design an improved score function of LDAs. The improved score function can handle the hesitancy of the linguistic terms. In light of the basic idea of Lin et al. (2021), we define an improved score function for LDAs as follows:

Definition 3. Let S be an LTS and $Ld = \left\{ \left(s^{(\tau)}, \beta^{(\tau)} \right) | s^{(\tau)} \in S, \beta^{(\tau)} > 0, \tau = 1, 2, \dots, |Ld| \right\}$ be an LDA on S, then an improved score function value of Ld is defined as:

$$Iscore(Ld) = \sum_{\tau=1}^{|Ld|} \left(1 - \mu \left(s^{(\tau)} \right) \right) \times \left(\sum_{\tau=1}^{|Ld|} \beta^{(\tau)} \Phi \left(s^{(\tau)} \right) \right), \tag{2}$$

where $\mu \left(s^{(\tau)} \right)$ is the hesitance measure of $s^{(\tau)}$ and it is computed based on (5):

$$\mu \left(s^{(\tau)} \right) = 1 - \frac{\left| \Phi(s^\tau) - \left(\sum\limits_{\tau=1}^{|Ld|} \beta^\tau \cdot \Phi(s^\tau) \right) \right|}{2\tau}. \tag{3}$$

From Definition 3, it can be observed that the $Iscore(Ld)$ considers the hesitance degree and uncertainty of LDAs.

When the input arguments of Model 2 are LDAs, according to the improved score function $Iscore(Ld)$, we modify it as follows:

Model 3

$$\theta_k = \min \frac{1 - \frac{\varepsilon}{n}\sum_{i=1}^{n}\left(\dfrac{\eta_{ik}^{-}}{\sum_{\tau=1}^{|Ld|}\left(1-\mu\left(s_{ik}^{(\tau)}\right)\right)\times\left(\sum_{\tau=1}^{|Ld|}\overline{\beta}_{ik}^{(\tau)}\Phi\left(s_{ik}^{(\tau)}\right)\right)}\right)}{1 + \frac{1}{h}\sum_{r=1}^{h}\left(\dfrac{\eta_{rD_k}^{+}}{\sum_{\tau=1}^{|Ld|}\left(1-\mu\left(s_{rk}^{(\tau)}\right)\right)\times\left(\sum_{\tau=1}^{|Ld|}\overline{\beta}_{rk}^{(\tau)}\Phi\left(s_{rk}^{(\tau)}\right)\right)}\right)}$$

$$s.t.\begin{cases} \sum_{j=1}^{m}\lambda_j\sum_{\tau=1}^{|Ld|}\left(1-\mu\left(s_{ij}^{(\tau)}\right)\right)\times\left(\sum_{\tau=1}^{|Ld|}\overline{\beta}_{ij}^{(\tau)}\Phi\left(s_{ij}^{(\tau)}\right)\right)+\eta_{ij}^{-}=\sum_{\tau=1}^{|Ld|}\left(1-\mu\left(s_{ik}^{(\tau)}\right)\right)\times\left(\sum_{\tau=1}^{|Ld|}\overline{\beta}_{ik}^{(\tau)}\Phi\left(s_{ik}^{(\tau)}\right)\right) \\ \sum_{j=1}^{m}\lambda_j\sum_{\tau=1}^{|Ld|}\left(1-\mu\left(s_{rDj}^{(\tau)}\right)\right)\times\left(\sum_{\tau=1}^{|Ld|}\overline{\beta}_{rDj}^{(\tau)}\Phi\left(s_{rDj}^{(\tau)}\right)\right)+\eta_{rD_k}^{-}=\sum_{\tau=1}^{|Ld|}\left(1-\mu\left(s_{rD_k}^{(\tau)}\right)\right)\times\left(\sum_{\tau=1}^{|Ld|}\overline{\beta}_{rD_k}^{(\tau)}\Phi\left(s_{rD_k}^{(\tau)}\right)\right) \\ \sum_{j=1}^{m}\lambda_j\sum_{\tau=1}^{|Ld|}\left(1-\mu\left(s_{rNDj}^{(\tau)}\right)\right)\times\left(\sum_{\tau=1}^{|Ld|}\overline{\beta}_{rNDj}^{(\tau)}\Phi\left(s_{rNDj}^{(\tau)}\right)\right)=\sum_{\tau=1}^{|Ld|}\left(1-\mu\left(s_{rND_k}^{(\tau)}\right)\right)\times\left(\sum_{\tau=1}^{|Ld|}\overline{\beta}_{rND_k}^{(\tau)}\Phi\left(s_{rND_k}^{(\tau)}\right)\right) \\ \eta_{ik}^{-}\geq 0,\ \eta_{rD_k}^{+}\geq 0,\ \lambda_j\geq 0(j=1,2,\ldots,m) \end{cases}$$

Let $Iscore_I(\Psi_{ik}) = \sum_{\tau}^{|Ld|}\left(1-\mu\left(s_{ik}^{(\tau)}\right)\right)\times\left(\sum_{\tau=1}^{|Ld|}\beta_{ik}^{(\tau)}\Phi\left(s_{ik}^{(\tau)}\right)\right)$ and $Iscore_O(\Psi_{rk}) = \sum_{\tau}^{|Ld|}\left(1-\mu\left(s_{rk}^{(\tau)}\right)\right)\times\left(\sum_{\tau=1}^{|Ld|}\beta_{rk}^{(\tau)}\Phi\left(s_{rk}^{(\tau)}\right)\right)$, then we simply the Model 3 as follows:

Model 4

$$\theta_k = \min \frac{1 - \frac{\varepsilon}{n}\sum_{i=1}^{n}\dfrac{\eta_{ik}^{-}}{Iscore_I(\overline{\Psi}_{ik})}}{1 + \frac{1}{h}\sum_{r=1}^{h}\dfrac{\eta_{rD_k}^{+}}{Iscore_O(\overline{\Psi}_{rD_k})}}$$

$$s.t.\begin{cases} \sum_{j=1}^{m}\lambda_j Iscore_I(\overline{\Psi}_{ij})+\eta_{ik}^{-}=Iscore_I(\overline{\Psi}_{ik}),\ (i=1,2,\ldots,n) \\ \sum_{j=1}^{m}\lambda_j Iscore_O(\overline{\Psi}_{rDj})-\eta_{rD_k}^{+}=Iscore_O(\overline{\Psi}_{rD_k}),\ (r=1,2,\ldots,h) \\ \sum_{j=1}^{m}\lambda_j Iscore_O(\overline{\Psi}_{rNDj})=Iscore_O(\overline{\Psi}_{rND_k}),\ (r=1,2,\ldots,z) \\ \eta_{ik}^{-}\geq 0,\ \eta_{rD_k}^{+}\geq 0,\ \lambda_j\geq 0, \end{cases}$$

where ε is a negligible positive integer; r^D and r^{ND} represent discretionary and non-discretionary outputs, respectively. η_{ik}^{-} and $\eta_{rD_k}^{+}$ are the slack variables connected with the inputs and the discretionary outputs, respectively; $Iscore(\overline{\Psi}_{ij}) = Iscore(\Psi_{ij}) - Iscore(\Psi_i)^{\min} > 0(\forall i, j)$; $Iscore(\overline{\Psi}_{rj}) = Iscore(\Psi_{rj}) - Iscore(\Psi_r)^{\min} > 0(\forall r, j)$; $Iscore_I(\overline{\Psi}_{ij})$ and $Iscore_O(\overline{\Psi}_{rj})$ are the score functions of the translated LDI and LDO respectively for the jth stadium, and $Iscore_O(\overline{\Psi}_{rk})$ are the score functions of the translated LDI and LDO for the kth stadium; λ_j is the weight assigned to the jth stadium. Thus, the optimal solution of Model 4 is $\left(\theta_k, \lambda_j^{*}, \eta_{ik}^{*-}, \eta_{rD_k}^{*+}\right)$.

The tourist satisfaction score θ_k is a means rather than an end to improving the service quality. The primary purpose of θ_k is to identify the pathway for each stadium to improve its service quality. If $\theta_k = 1$ then stadium k is considered the best practice; otherwise, it is an underperformed stadium. It must be emphasized that the condition $\theta_k = 1$ is equal to both $\eta_{ik}^{*-} = 0$ and $\eta_{rD_k}^{*+} = 0$ (Tone et al., 2020). As a result, the stadia can be divided

into two classes, i.e., benchmarks (best-practice stadia) if a stadium obtains a score of 1 and underperformed stadia needing a benchmark for service quality improvement. To improve the tourist satisfaction of an inefficient stadium, we determine its benchmark as follows:

$$\Lambda_k = \left\{ j | \lambda_j^* > 0 \right\} (j = 1, 2, \ldots, n). \tag{4}$$

When a stadium underperforms, the output level in (4) can be employed as the origin for setting benchmarks to enhance tourist satisfaction. Each underperformed stadium can become the best practice stadium by fine-tuning its operations to the benchmark goals determined by the best practice stadium that describes its reference frontier. An inefficient stadium k can improve and become efficient by increasing the output shortage as follows:

$$\widehat{y}_{rk} = Iscore_O \left(\overline{\Psi}_{rD_k} \right) + \eta_{rD_k}^{+*}. \tag{5}$$

3.6 Service quality association

In this subsection, we re-investigate the sentiment orientations for an underperformed stadium to estimate the quality association between the service quality attributes for improvement. This step utilizes association rule mining (Agrawal et al., 1993) to perform the association analysis. The association rule explains the similarity (or dissimilarity) between two service quality attributes of the underperforming stadium. Let $F_r \rightarrow F_v$ be the association rule between the service quality attributes F_r and F_v of each stadium ($r, v = 1, 2, \ldots, q; r \neq v$). According to the association rule, we compute the support, confidence, and lift of the rule $F_r \rightarrow F_v$ as follows (Alam et al., 2021):

$$support(F_r \rightarrow F_v) = \frac{|F_r \cup F_v|}{|q|} = supp(F_r \cup F_v), \tag{6}$$

$$conf(F_r \rightarrow F_v) = \frac{Supp(F_r \cup F_v)}{Supp(F_r)}, \tag{7}$$

$$lift(F_r \rightarrow F_v) = \frac{Supp(F_r \rightarrow F_v)}{Supp(F_r) \cdot Supp(F_v)}. \tag{8}$$

Support indicates the frequency of combining two key service attributes in the collection of service attributes. Confidence indicates the normalized effect of a key service feature on a rule and can be explained as the credibility of one key service feature impacting another feature. The lift is the best option to determine the polarity of association rules since the support and confidence cannot achieve that. According to each association rule $F_r \rightarrow F_v$, the following remarks can be concluded:

Remark 1. If $lift(F_r \rightarrow F_v) = 1$, then F_r, and F_v are autonomous of each other.

Remark 2. If $lift(F_r \rightarrow F_v) > 1$, then F_r, and F_v are positively interrelated.

Remark 3. If $lift(F_r \rightarrow F_v) < 1$, then F_r and F_v are negatively interrelated.

Referring to the works of Xu et al. (2019) and Liang et al. (2020), we use the centrality to calculate the weight of the service attributes. The weight importance is computed based on the weight of the association rule ϖ_{rv}, which is obtained according to the following remarks:

Remark 4. If $lift(F_r \rightarrow F_v) \geq 1$, then ϖ_{rv} is $\varpi_{rv} = conf(F_r \rightarrow F_v)$.

Remark 5. If $lift(F_r \rightarrow F_v) < 1$, then $\varpi_{rv} = -conf(F_r \rightarrow F_v)$.

The degree of centrality Ω_r sums the outdegree of F_r and the indegree of F_r together. The outdegree value F_r is determined by adding the weights of all association rules that go to this key service feature. In contrast, the indegree value F_r is determined by adding the weights of all association rules that go out of this key service feature. Therefore, the degree of centrality of the key service feature F_r is ascertained as follows:

$$\Omega(F_r) = \text{Outdegree}(F_r) + \text{Indegree}(F_r). \qquad (9)$$

According to the degree of centrality, we determine the weight and importance of the key service attributes as follows:

$$w(F_r) = \frac{\Omega(F_r)}{\sum_{i=1}^{z}(F_r)}. \qquad (10)$$

Based on the weight information, the service quality attribute with the highest weight is the most important. The lift values of the association mining indicate the direction of the relationship between the attributes. Therefore, we can deduce the improvement pathway by utilizing the weight of the attributes and their lift values. The improvement algorithm is presented as follows:

Step 1 Arrange each underperforming stadium's service quality attributes in a descending order based on Eq. (10).

Step 2 Choose the first service quality attribute for further analysis.

Step 3 Check the association rule between the first-ranked attribute and the other attributes that need improvement.

Step 4 Based on Remarks 1–4, the direction of the association between the first-ranked attribute and the other attributes can be ascertained.

(a) If the lift values are greater than 1, improving the first-ranked attribute can translate into the growth of the other attributes. Hence, the underperforming stadium can follow this path.
(b) If the lift values are less than 1, improving the first-ranked attribute can reduce the growth of the other attributes. Hence, the underperforming stadium should ignore this path.
(c) If the lift values are equal to 1, the improvement of the first-ranked attribute does not affect the other attributes. Hence, the attributes should be improved individually.

4 Decision analysis

In this study, we experiment using the Union of European Football Associations (UEFA) Elite stadia. UEFA categorizes stadia as category one, two, three, or UEFA elite stadia (worldstadiumdatabase.com). A stadium must be recognized as an elite stadium to host the final of the UEFA Champions League or the UEFA Europa League, UEFA's two elite club competitions. These stadia are adding new usages within stadia to attract diverse tourists throughout the year (Edensor et al., 2021), with some moving to new multi-functional arenas. These tourists seek experiences other than watching football matches.

First, we sample the top seven football leagues (England, Spain, Italy, Germany, France, Portugal, and the Netherlands) according to UEFA's 2020/2021 country coefficients. Then, the ten most popular elite stadia in these leagues with a capacity of 50,000–100,000 are

selected for further studies. This study draws on TripAdvisor reviews to comprehensively investigate tourist satisfaction within a football stadium. We crawl 15,177 online reviews, including ratings and text reviews, using python programming based on selenium. The data is collected from 2018 to 2020, and the sample includes only reviews written in English.

4.1 Data acquisition

According to Sect. 3.2.2, the textual reviews are preprocessed using RegEx, SymSpell, and SpaCy natural language techniques. The preprocess allows us to have fine-grain reviews to extract the key service attributes. Employing the processes of Sect. 3.2.3, we identify the key service quality attributes from the cleaned textual reviews. Five topics are obtained from the latent Dirichlet allocation model. The topics and their keywords are presented in Table 1.

The study compares the results, and conflicts are handled through group discussions to arrive at the final description of each topic, as suggested by Glowacki et al. (2018). Also, we refer to the textual reviews to obtain insight into renaming each topic. Hence, the following five service quality attributes (topics) are ascertained: facilities (F_1), service (F_2), atmosphere (F_3), guest support (F_4), and location (F_5).

The Latent Dirichlet allocation model produces the dominant topic for each textual review. Then, we can group similar textual reviews under each topic (service quality attribute.). The VADER sentiment algorithm is applied to obtain the sentiment scores of the textual reviews under each service quality attribute. Based on the translation rules of Sect. 3.3, we convert the sentiment scores into LDAs. Also, the ratings are translated as well. Subsequently, we build the input and output matrix in Table 2. In Table 2, the LDAs represent the evaluation of each stadium concerning each variable. For example, the LDAs of facilities concerning Allianz Arena denote that 89.5% of the tourists have very positive opinions about the facilities while 7% have positive opinions. Also, 1% of the tourists have general opinions. Meanwhile, 1% of the tourists have negative opinions about the facilities while 0.6% have very negative opinions.

4.2 Benchmarking analysis results

Using Model 3 of subsection 3.5, we derive benchmarking results of the linguistic distribution output-oriented non-discretionary BP-SBM method. The results are displayed in Table 3. In Table 3, the scores of the LDAs are ascertained using the improved score function $Iscore(Ld)$ and (2).

A stadium with a score of 1 has the highest grade of tourist satisfaction and has the opportunity to become a target for an underperformed stadium having a score lower than 1. Table 3 shows that Camp Nou, San Siro, Stade de France, Veltins-Arena, Stadion Feijenoord, and Old Trafford are the best practice stadia. Allianz Arena, Santiago Bernabeu, Stadio Olimpico, and Wembley are the underperformed stadia. Accordingly, the underperforming stadia choose stadium numbers 3, 6, 7, and 10 as their benchmarks. Camp Nou acts as a benchmark for four underperformed stadia, i.e., Allianz Arena, Santiago Bernabeu, Stadio Olimpico, and Wembley. San Siro is a benchmark for two underperformed stadia, i.e., Allianz Arena and Stadio Olimpico. Also, Stade de France is a benchmark for only one stadium, i.e., Wembley. Old Trafford is a benchmark for four stadia, i.e., Allianz Arena, Santiago Bernabeu, Stadio Olimpico, and Wembley. On the one hand, Camp Nou and Old Trafford are the most selected benchmarks for the underperforming stadia. On the other hand, Veltins-Arena and Stadion Feijenoord are ignored as benchmarks by the underperformed stadia though they

Table 1 Topics and related words

Topics	Related keywords		
Topic 1	Museum	Pitch	Event
	Football	Room	Game
	Player	Concert	Photo
	Press		
Topic 2	Fan	Guide	Ticket
	Way	Time	People
	Help	Olympics	Man
	New		
Topic 3	Fan	Day	Place
	Experience	Great	Press
	Visit	Old	Area
	Atmosphere		
Topic 4	Staff	Legend	Interest
	Team	Ground	Information
	Guide	support	Thank
	Organize		
Topic 5	Food	Match	Birthday

Table 1 (continued)

Topics	Related keywords		
	Seat	Venue	Unit
	Drink	View	West
	Park	brilliant	

Table 2 Linguistic distribution assessments based on the sentiment scores

No	Stadium	Input	Outputs				
		Stadium reputation	Facilities	Service	Atmosphere	Guest support	Location
1	Allianz Arena	$s_{-2}\{0.003\}$, $s_{-1}\{0.005\}$, $s_0\{0.013\}$, $s_1\{0.038\}$, $s_2\{0.943\}$	$s_{-2}\{0.006\}$, $s_{-1}\{0.011\}$, $s_0\{0.011\}$, $s_1\{0.077\}$, $s_2\{0.895\}$	$s_{-2}\{0.033\}$, $s_{-1}\{0.037\}$, $s_0\{0.011\}$, $s_1\{0.056\}$, $s_2\{0.862\}$	$s_{-2}\{0.041\}$, $s_{-1}\{0.050\}$, $s_0\{0.023\}$, $s_1\{0.123\}$, $s_2\{0.763\}$	$s_{-2}\{0.005\}$, $s_{-1}\{0.032\}$, $s_0\{0.020\}$, $s_1\{0.071\}$, $s_2\{0.873\}$	$s_{-2}\{0.033\}$, $s_{-1}\{0.066\}$, $s_1\{0.088\}$, $s_2\{0.812\}$
2	Santiago Bernabeu	$s_{-2}\{0.018\}$, $s_{-1}\{0.009\}$, $s_0\{0.025\}$, $s_1\{0.096\}$, $s_2\{0.851\}$	$s_{-2}\{0.009\}$, $s_{-1}\{0.019\}$, $s_0\{0.009\}$, $s_1\{0.046\}$, $s_2\{0.917\}$	$s_{-2}\{0.008\}$, $s_{-1}\{0.016\}$, $s_0\{0.008\}$, $s_1\{0.035\}$, $s_2\{0.932\}$	$s_{-2}\{0.111\}$, $s_{-1}\{0.063\}$, $s_0\{0.032\}$, $s_2\{0.794\}$	$s_{-2}\{0.013\}$, $s_{-1}\{0.011\}$, $s_0\{0.013\}$, $s_1\{0.066\}$, $s_2\{0.897\}$	$s_{-2}\{0.027\}$, $s_{-1}\{0.017\}$, $s_0\{0.014\}$, $s_1\{0.065\}$, $s_2\{0.878\}$
3	Camp Nou	$s_{-2}\{0.120\}$, $s_{-1}\{0.010\}$, $s_0\{0.038\}$, $s_1\{0.135\}$, $s_2\{0.698\}$	$s_{-2}\{0.006\}$, $s_{-1}\{0.011\}$, $s_0\{0.006\}$, $s_1\{0.086\}$, $s_2\{0.891\}$	$s_{-2}\{0.031\}$, $s_{-1}\{0.020\}$, $s_0\{0.004\}$, $s_1\{0.063\}$, $s_2\{0.882\}$	$s_{-2}\{0.181\}$, $s_{-1}\{0.117\}$, $s_0\{0.011\}$, $s_1\{0.064\}$, $s_2\{0.628\}$	$s_{-2}\{0.023\}$, $s_{-1}\{0.026\}$, $s_0\{0.004\}$, $s_1\{0.105\}$, $s_2\{0.842\}$	$s_{-2}\{0.079\}$, $s_{-1}\{0.066\}$, $s_0\{0.029\}$, $s_1\{0.104\}$, $s_2\{0.722\}$
4	Stadio Olimpico	$s_{-2}\{0.017\}$, $s_{-1}\{0.001\}$, $s_1\{0.230\}$, $s_2\{0.752\}$	$s_{-2}\{0.019\}$, $s_0\{0.038\}$, $s_1\{0.019\}$, $s_2\{0.925\}$	$s_{-1}\{0.027\}$, $s_0\{0.005\}$, $s_1\{0.075\}$, $s_2\{0.892\}$	$s_{-2}\{0.167\}$, $s_{-1}\{0.083\}$, $s_1\{0.083\}$, $s_2\{0.667\}$	$s_{-2}\{0.022\}$, $s_{-1}\{0.018\}$, $s_1\{0.031\}$, $s_2\{0.928\}$	$s_{-2}\{0.040\}$, $s_{-1}\{0.060\}$, $s_0\{0.040\}$, $s_1\{0.060\}$, $s_2\{0.800\}$

Table 2 (continued)

No	Stadium	Input	Outputs					
		Stadium reputation	Facilities	Service	Atmosphere	Guest support	Location
5	Wembley Stadium	$s_0\{0.123\}$, $s_2\{0.877\}$	$s_1\{0.039\}$, $s_2\{0.961\}$	$s_{-2}\{0.006\}$, $s_{-1}\{0.007\}$, $s_0\{0.004\}$, $s_1\{0.031\}$, $s_2\{0.952\}$	$s_{-2}\{0.053\}$, $s_{-1}\{0.092\}$, $s_1\{0.079\}$, $s_2\{0.776\}$	$s_{-2}\{0.019\}$, $s_{-1}\{0.014\}$, $s_0\{0.005\}$, $s_1\{0.023\}$, $s_2\{0.939\}$	$s_{-2}\{0.022\}$, $s_{-1}\{0.034\}$, $s_0\{0.011\}$, $s_1\{0.034\}$, $s_2\{0.899\}$
6	San Siro	$s_{-2}\{0.024\}$, $s_{-1}\{0.014\}$, $s_0\{0.037\}$, $s_1\{0.024\}$, $s_2\{0.901\}$	$s_{-1}\{0.056\}$, $s_2\{0.944\}$	$s_{-2}\{0.029\}$, $s_{-1}\{0.029\}$, $s_0\{0.017\}$, $s_1\{0.052\}$, $s_2\{0.873\}$	$s_{-2}\{0.121\}$, $s_{-1}\{0.061\}$, $s_0\{0.030\}$, $s_1\{0.091\}$, $s_2\{0.697\}$	$s_{-2}\{0.024\}$, $s_{-1}\{0.059\}$, $s_1\{0.118\}$, $s_2\{0.800\}$	$s_{-2}\{0.075\}$, $s_{-1}\{0.075\}$, $s_0\{0.100\}$, $s_1\{0.050\}$, $s_2\{0.700\}$
7	Stade de France	$s_{-2}\{0.034\}$, $s_{-1}\{0.024\}$, $s_0\{0.034\}$, $s_1\{0.135\}$, $s_2\{0.773\}$	$s_{-1}\{0.013\}$, $s_0\{0.006\}$, $s_1\{0.022\}$, $s_2\{0.959\}$	$s_{-2}\{0.013\}$, $s_{-1}\{0.011\}$, $s_0\{0.007\}$, $s_1\{0.033\}$, $s_2\{0.937\}$	$s_{-2}\{0.161\}$, $s_{-1}\{0.016\}$, $s_1\{0.129\}$, $s_2\{0.694\}$	$s_{-2}\{0.014\}$, $s_{-1}\{0.014\}$, $s_1\{0.036\}$, $s_2\{0.935\}$	$s_{-2}\{0.012\}$, $s_{-1}\{0.041\}$, $s_0\{0.012\}$, $s_1\{0.094\}$, $s_2\{0.842\}$
8	Veltins Arena	$s_{-2}\{0.100\}$, $s_2\{0.900\}$	$s_{-2}\{0.005\}$, $s_{-1}\{0.005\}$, $s_1\{0.011\}$, $s_2\{0.979\}$	$s_{-2}\{0.002\}$, $s_{-1}\{0.007\}$, $s_0\{0.002\}$, $s_1\{0.027\}$, $s_2\{0.961\}$	$s_{-2}\{0.074\}$, $s_2\{0.926\}$	$s_{-2}\{0.014\}$, $s_1\{0.041\}$, $s_2\{0.945\}$	$s_{-2}\{0.047\}$, $s_{-1}\{0.063\}$, $s_1\{0.078\}$, $s_2\{0.813\}$

Table 2 (continued)

No	Stadium	Input	Outputs				
		Stadium reputation	Facilities	Service	Atmosphere	Guest support	Location
9	Stadium Feijenoord	$s_{-2}\{0.003\}$, $s_{-1}\{0.005\}$, $s_0\{0.013\}$, $s_1\{0.038\}$, $s_2\{0.943\}$	$s_{-1}\{0.008\}$, $s_0\{0.008\}$, $s_1\{0.016\}$, $s_2\{0.968\}$	$s_1\{0.022\}$, $s_2\{0.978\}$	$s_{-2}\{0.091\}$, $s_2\{0.909\}$	$s_2\{1.000\}$	$s_{-2}\{0.040\}$, $s_{-1}\{0.080\}$, $s_2\{0.880\}$
10	Old Trafford	$s_{-2}\{0.001\}$, $s_{-1}\{0.070\}$, $s_1\{0.010\}$, $s_2\{0.919\}$	$s_2\{1.000\}$	$s_{-2}\{0.041\}$, $s_{-1}\{0.050\}$, $s_0\{0.023\}$, $s_1\{0.123\}$, $s_2\{0.763\}$	$s_{-2}\{0.007\}$, $s_{-1}\{0.013\}$, $s_0\{0.010\}$, $s_1\{0.059\}$, $s_2\{0.912\}$	$s_{-1}\{0.008\}$, $s_0\{0.008\}$, $s_1\{0.016\}$, $s_2\{0.968\}$	$s_{-2}\{0.045\}$, $s_{-1}\{0.057\}$, $s_0\{0.007\}$, $s_1\{0.102\}$, $s_2\{0.789\}$

Table 3 Tourist satisfaction scores, LDAs scores, and benchmarking goals

No.	Stadia	Scores	Scores of the LDAs				Benchmarking goals				Benchmarks (Stadium No.)
			Facilities	Service	Atmosphere	Guest support	Facilities	Service	Atmosphere	Guest support	
1	Allianz Arena	0.822	4.394	4.138	3.786	3.243	5.846	5.132	4.896	3.243	3,6,10
2	Santiago Bernabeu	0.765	4.390	4.315	4.481	2.672	6.641	5.997	5.802	2.762	3,10
3	Camp Nou	1.000	4.398	3.928	4.029	1.438	4.398	3.928	4.029	1.438	3
4	Stadio Olimpico	0.794	4.345	4.323	4.354	1.750	5.324	4.806	4.354	2.973	3,6,10
5	Wembley	0.864	4.845	4.413	4.669	2.973	6.045	5.570	5.225	2.973	3,7,10
6	San Siro	1.000	4.349	3.557	3.907	2.230	4.349	3.557	3.907	2.230	6
7	Stade de France	1.000	4.712	4.470	4.493	2.222	4.712	4.470	4.493	2.222	7
8	Veltins-Arena	1.000	4.818	4.619	4.755	3.882	4.818	4.619	4.755	3.888	8
9	Stadion Feijenoord	1.000	4.778	5.000	4.912	3.643	4.778	5.000	4.912	3.643	9
10	Old Trafford	1.000	5.000	4.778	3.243	4.440	5.000	4.778	3.243	4.440	10

Table 4 Improvement goals of the underperformed stadia

No	Stadium	Facilities	Service	Atmosphere	Guest support
1	Allianz Arena	1.452	0.994	1.110	0
2	Santiago Bernabeu	2.251	1.682	1.321	0.090
4	Stadio Olimpico	0.979	0.483	0	1.223
5	Wembley	1.200	1.157	0.556	0

are best practice stadia. In line with the benchmarking goals, the underperformed stadia can achieve tourists satisfaction score of 1 and be the best practices as their benchmarks.

Based on the scores of the LDAs and the benchmarking goals, we can deduce the improvement values for the underperformed stadia. The outcome is given in Table 4.

According to Table 4, the underperforming stadia can improve their outputs (service quality attributes) through the improvement goals. Allianz Arena obtained satisfaction of 0.822. Allianz Arena can become a best practice stadium as its benchmark by improving on the following three service quality attributes: facilities, service, and atmosphere. Santiago Bernabeu obtained the lowest satisfaction of 0.765. Santiago Bernabeu can become the best practical stadium by improving on the following four service quality attributes: facilities, service, atmosphere, and guest support. The satisfaction for Stadio Olimpico is 0.794. Stadio Olimpico needs to improve on three service attributes: facilities, service, and guest support. For Wembley, the satisfaction is 0.864. Wembley qualifies as a best practice stadium by improving three service quality attributes: facilities, service, and atmosphere.

4.3 Service quality association results

The benchmarking improvement goals convey the service quality attributes needed for improvement. However, due to limited resources, we need to know which one is more relevant to improve first and how its improvement can ensure the growth of the other attributes. Utilizing XLSTAT software (XLSTAT, 2007), we deduce some association rules of the service quality attributes to help an underperformed stadium.

According to (6–8), we compute the support, confidence, and lift measures. We set the minimum threshold for support to 0.1 and the minimum threshold for confidence to 0.05. Results for the support, the confidence, and the lift of the four underperforming stadia are presented in Table 5.

Using (10), we compute the relative importance of the service quality attributes for each underperformed stadium. The result is listed in Table 6. The service quality attribute with the highest weight importance score for each underperforming stadium is selected as the pivotal point for the improvement analysis.

Based on Steps 1–4 in subsection 3.6, we provide some improvement paths for the underperformed stadia. In Table 6 F_1 is the most important service quality attribute for Allianz Arena. Since the lift values (1.155, 1.156, and 1.620) are greater than 1, we know that F_1 has a positive association with F_2, F_3 and F_4. This means that the improvement of F_1 can enhance the other the service quality attributes. For Santiago Bernabeu, F_1 is the most important attribute for improvement. According to the lift values (1.002, 3.375, and 1.620) for this stadium, F_1 has a positive association with F_2, F_3, and F_4. Hence, the improvement of F_1 can drive the growth of the other attributes. F_1 is the most important service attribute for

Table 5 Association rules for the underperformed stadia

Rule	Allianz Arena			Santiago Bernabeu			Stadio Olimpico			Wembley		
	Support	Confidence	Lift	Support	Confidence	Lift	Support	Confidence	Lift	Support	Confidence	Lift
$F_1 \rightarrow F_2$	0.642	0.753	1.155	0.287	0.986	1.002	0.229	1.000	1.205	0.262	0.994	0.999
$F_1 \rightarrow F_3$	0.517	0.606	1.156	0.285	0.981	3.375	0.229	1.000	1.429	0.261	0.989	1.540
$F_1 \rightarrow F_4$	0.836	0.980	1.620	0.288	0.991	1.620	0.229	1.000	1.231	0.262	0.994	3.199
$F_2 \rightarrow F_1$	0.642	0.985	1.155	0.287	0.289	1.002	0.229	0.276	1.205	0.262	0.264	0.999
$F_2 \rightarrow F_3$	0.517	0.793	1.512	0.287	0.289	0.994	0.695	0.838	1.198	0.641	0.644	1.002
$F_2 \rightarrow F_4$	0.640	0.981	0.994	0.606	0.611	1.714	0.830	1.000	0.994	0.309	0.311	2.000
$F_3 \rightarrow F_1$	0.517	0.986	1.156	0.285	0.981	3.375	1.429	0.327	3.375	0.261	0.406	1.540
$F_3 \rightarrow F_2$	0.517	0.986	1.156	0.287	0.986	0.944	1.198	0.994	0.944	0.641	0.998	1.002
$F_3 \rightarrow F_4$	0.512	0.977	1.231	0.288	0.991	1.620	1.321	1.000	1.620	0.308	0.479	1.542
$F_4 \rightarrow F_1$	0.836	0.853	1.620	0.288	0.471	1.620	0.231	0.229	1.620	0.262	0.844	3.199
$F_4 \rightarrow F_2$	0.640	0.653	0.994	0.606	0.991	1.714	0.994	0.830	1.714	0.309	0.995	2.000
$F_4 \rightarrow F_3$	0.512	0.523	1.231	0.288	0.471	1.620	1.321	0.700	1.620	0.308	0.991	1.542

Table 6 Weight importance of service quality attributes for the underperformed stadia

No	Stadium	F_1	F_2	F_3	F_4
1	Allianz Arena	0.379	0.138	0.358	0.125
2	Santiago Bernabeu	0.362	0.123	0.166	0.349
4	Stadio Olimpico	0.368	0.091	0.276	0.265
5	Wembley	0.154	0.132	0.352	0.361

Stadio Olimpico. From the lift values (1.205, 1.429, and 1.231) F_1 has a positive relationship with F_2, F_3, and F_4. Therefore, the improvement of F_1 enhances the other attributes. For Wembley, F_4 is the most important service quality attribute for improvement. In line with the lift values (3.199, 2.000, and 1.542), F_4 has a positive association with F_1, F_2 and F_3. Therefore, the growth of F_4 improves the other attributes.

4.4 Comparative analysis

In what follows, we conduct a comparative analysis using the Tripadvisor.com rating survey system. Tripadvisor.com provides an overall rating for stadia, representing the degree of tourists' satisfaction. The rating survey system usually generates tourists' satisfaction scores ranging from 1 to 5, with 5 being the highest. Based on the overall ratings, we can rank the respective stadia. As a result, we compare our degree of tourist satisfaction obtained from the proposed model with Tripadvisor.com. Figure 2 shows the result of the comparison. We observe that the rating scores have little discrimination power over the stadia. From Fig. 2, 8 of the 10 stadia have a rating score of 4.5. For instance, stadia 1, 2, 3, and 5 have a similar
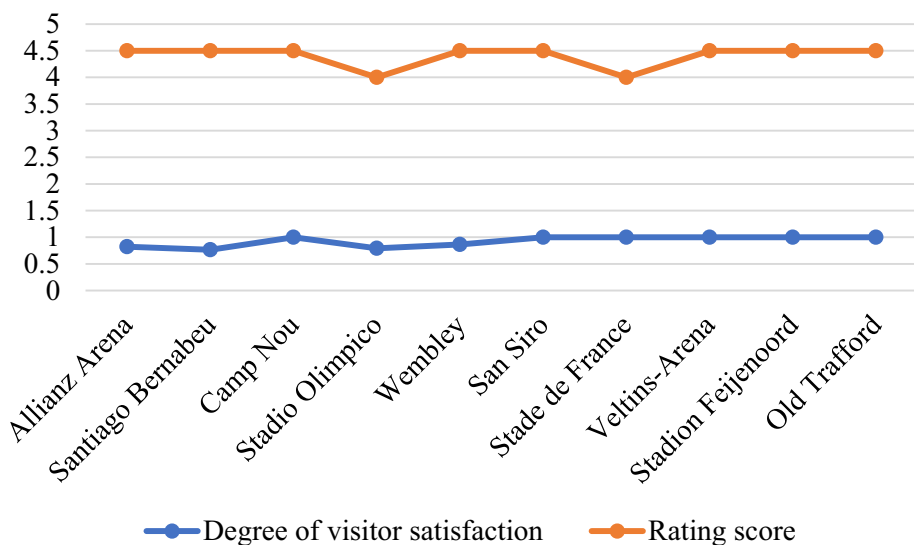
Fig. 2 Comparative degrees of tourist satisfaction and the rating scores estimated by Tripadvisor.com

rating score of 4.5; however, their estimated degrees of tourist satisfaction obtained from the proposed model are 0.822, 0.765, 1, and 0.864, respectively. Extant studies (López Fernández & Serrano Bedia, 2004; Núñez-Serrano et al., 2014; Oliveira et al., 2013) have raised issues concerning the worth of rating scores as a good estimator of satisfaction. They concluded that the rating survey method is irrelevant in measuring the effectiveness of decision-making units.

We conduct another experiment to examine tourist satisfaction considering the discretionary and non-discretionary nature of the service quality attributes. Our proposed model contends that one of service quality's attributes (i.e., location) is non-discretionary because decision-makers have little control over it. We consider the situation where all the service quality attributes are discretionary and compare them with our results. The experimental outcomes are fitted in Fig. 3.

Figure 3 reveals that the two situations under observation produce different tourist satisfaction degrees. Considering that all the key service quality attributes are discretionary, stadia 3, 7, and 10 (i.e., Camp Nou, Stade de France, and Old Trafford) are the best practice stadia, while the rest are underperformed. While our proposed case (non-discretionary) considers four stadia to be underperformed, the other case (discretionary) selects seven stadia. The Pearson correlation coefficient estimated between these two methods is 0.354. This implies that the two methods are marginally associated yet necessarily distinct to contribute to the stadium benchmarking. Selecting service quality attributes as outputs for the SBM DEA is crucial as it affects the degrees of tourist satisfaction. Hence, decision-makers should consider the types of outputs, i.e., discretionary and non-discretionary, when evaluating customer satisfaction. We contend that our proposed model, which considers non-discretionary outputs, is superior because it models real-world problems where decision-makers do not always control all their resources.
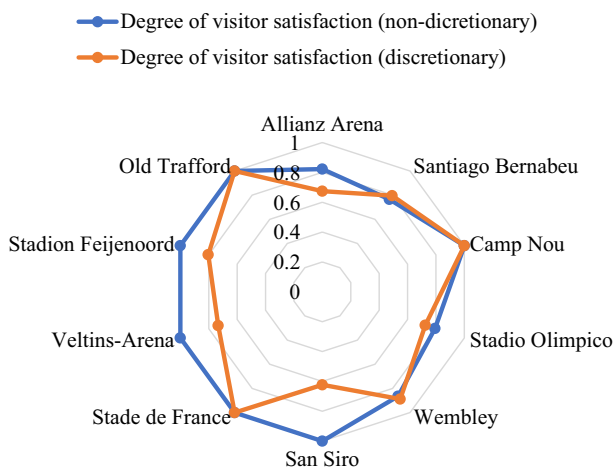

Fig. 3 Comparative analysis considering discretionary and non-discretionary outputs in the proposed SBM DEA

5 Implications

This paper develops a linguistic distribution SBM of DEA based on online reviews to model the degree of tourist satisfaction regarding football stadium service quality. The results provide some practical implications in sports tourism benchmarking and offer insights for decision-makers. Theoretically, this study extends previous research on sports tourism by designing data-driven decision support for benchmarking football stadia. Although tourists' experiences play a significant role in improving football facilities, previous studies focused much on supply-sided perspectives. Therefore, this study shows that analyzing tourist experiences through online reviews can be valuable for investigating football stadium service quality. The results shed light on how text mining approaches can extract service quality attributes for benchmarking tourist satisfaction.

Additional value is added to this research in that it analyzes tourists' sentiments through the concept of LDAs. LDAs enhance the applicability of tourist opinions by modeling the uncertainties associated with them. Through LDAs, we can effortlessly address the problem of group evaluation. This study also contributes by improving the SBM of DEA to accommodate LDAs. The proposed linguistic distribution SBM of DEA is a powerful method for measuring decision-making units' performance. Moreover, the improvement pathway of quality association designed in this study enhances benchmarking analysis.

Also, the study provides practical implications for the sports industry, notably football stadium tourism. Stadium managers can leverage the data mining techniques discussed in this study to identify significant service quality attributes from the voluminous online reviews. The sentiment analysis is helpful for managers to discover tourists' feelings and perceptions regarding service quality attributes. As a result, managers may recognize the strength and weaknesses of their sports facilities. The benchmarking studies will aid decision-makers in gauging the degrees of tourist satisfaction (benchmarking scores) and develop strategies to reach the status of best-practice stadia. Also, through the association analysis, managers can ascertain the importance of service quality attributes in improving tourist satisfaction.

6 Conclusion

The past decade has seen top-flight football teams positioning themselves as tourist attractions, influencing tourists to visit their stadia attractions (e.g., live matches, stadium tours, and museums) on match days or non-match days. Consequently, a series of investigations emerged to cover this new tourism trend through management, tourism, heritage, geography, and economic lenses. This study focuses on stadium tours other than involvement in sports activities. Measuring tourists' satisfaction through service quality is necessary to improve stadium tourism. Therefore, this study proposes a decision-support model to measure the extent of tourist satisfaction and define benchmarking plans for football stadia utilizing online reviews. We employ text mining techniques such as Latent Dirichlet allocation to extract service quality attributes according to the proposed model. Then, we conduct sentiment analysis to reveal the polarities of the text reviews. To solve uncertainty and fuzziness in online reviews, we employ LDAs to transform the sentiment opinions. In addition, we design an improved score function for LDAs and integrate it with an output-oriented non-discretionary BP-SBM of DEA to benchmark the degree of tourist satisfaction of football stadia tours. Meanwhile, utilizing association rule mining, we examine the interrelationship between the

service quality attributes of an underperformed stadium. Based on the strength of the association pattern, we design an approach for improving the service quality attributes. The strength of the proposed decision-support model is that transparent processes are followed to identify the data for the DEA model. Instead of relying on traditional service quality models and surveys, the proposed model builds a service quality framework from online reviews, which are timely and less costly. Also, the introduction of LDAs aid in analyzing group evaluations devoid of uncertainties. Again, the proposed model goes beyond just benchmarking tourist satisfaction to provide improvement pathways for the underperforming stadium where less attention has been given. The limitation of this study is that it measures tourist satisfaction for a specific period instead of overtime periods. Therefore, future studies would upgrade the proposed model using a dynamic DEA model to examine the progressive enhancement of tourist satisfaction over time.

References

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*. https://doi.org/10.1145/170036.170072

Alam, T. M., Shaukat, K., Hameed, I. A., Khan, W. A., Sarwar, M. U., Iqbal, F., & Luo, S. (2021). A novel framework for prognostic factors identification of malignant mesothelioma through association rule mining. *Biomedical Signal Processing and Control*. https://doi.org/10.1016/j.bspc.2021.102726

Álvarez-García, J., González-Vázquez, E., del de la Río-Rama, M. C., & Durán-Sánchez, A. (2019). Quality in customer service and its relationship with satisfaction: An innovation and competitiveness tool in sport and health centers. *International Journal of Environmental Research and Public Health*. https://doi.org/10.3390/ijerph16203942

Anderson, E. W., & Sullivan, M. W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing Science, 12*(2), 125–143. https://doi.org/10.1287/MKSC.12.2.125

Andreassen, T. W., & Lindestad, B. (1998). Customer loyalty and complex services. The impact of corporate image on quality, customer satisfaction and loyalty for customers with varying degrees of service expertise. *International Journal of Service Industry Management*. https://doi.org/10.1108/09564239810199923

Aniket, K., & Roshan, D. (2021). *Sports Tourism Market Size, Share\Industry analysis: 2030*. https://www.alliedmarketresearch.com/sports-tourism-market-A13076

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*. https://doi.org/10.1287/mnsc.30.9.1078

Blanchard, R. F., & Galloway, R. L. (1994). Quality in retail banking. *International Journal of Service Industry Management, 5*(4), 5–23. https://doi.org/10.1108/09564239410068670/FULL/XML

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 4*, 993–1022. https://doi.org/10.1016/b978-0-12-411519-4.00006-9

Bloemer, J., de Ruyter, K., & Peeters, P. (1998). Investigating drivers of bank loyalty: The complex relationship between image, service quality and satisfaction. *International Journal of Bank Marketing*. https://doi.org/10.1108/02652329810245984

Bonser, S., Hughes, B. R., & Calautit, J. K. (2020). Investigation of the impact of roof configurations on the wind and thermal environment in football stadiums in hot climates. *International Journal of Ventilation*. https://doi.org/10.1080/14733315.2019.1665861

Bunds, K. S., Casper, J. M., Frey, H. C., & Barrett, M. (2019). Air pollution at college football games: Developing a methodology for measuring air pollutant exposure in a sport event microenvironment. *Event Management*. https://doi.org/10.3727/152599518X15403853721484

Chalip, L., & McGuirty, J. (2004). Bundling sport events with the host destination. *Journal of Sport and Tourism*. https://doi.org/10.1080/1477508042000320241

Chen, H., & Rahman, I. (2018). Cultural tourism: An analysis of engagement, cultural contact, memorable tourism experience and destination loyalty. *Tourism Management Perspectives*. https://doi.org/10.1016/j.tmp.2017.10.006

Chiles, B. (2021). Shrouded prices and firm reputation: Evidence from the U.S. Hotel Industry. *Management Science*. https://doi.org/10.1287/mnsc.2019.3566

Cho, H., Joo, D., & Chi, C. G. (2019). Examining nostalgia in sport tourism: The case of US college football fans. *Tourism Management Perspectives*. https://doi.org/10.1016/j.tmp.2018.11.002

Darden, W. R., & Schwinghammer, J. K. (1985). The influence of social characteristics on perceived quality in patronage choice behavior. *Perceived Quality*, 161–172.

Darko, A. P., & Liang, D. (2022a). A heterogeneous opinion-driven decision-support model for tourists' selection with different travel needs in online reviews. *Journal of the Operational Research Society, 0*, 1–18. https://doi.org/10.1080/01605682.2022.2035274

Darko, A. P., & Liang, D. (2022b). Modeling customer satisfaction through online reviews: A FlowSort group decision model under probabilistic linguistic settings. *Expert Systems with Applications, 195*(February), 116649. https://doi.org/10.1016/j.eswa.2022.116649

de Matta, R., & Lowe, T. J. (2021). Product price alignment with seller service rating and consumer satisfaction. *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-04318-3

Edensor, T., Millington, S., Steadman, C., & Taecharungroj, V. (2021). Towards a comprehensive understanding of football stadium tourism. *Journal of Sport and Tourism*. https://doi.org/10.1080/14775085.2021.1884589

Ryan Erskine. (2018). *Study: 97% of Business Owners Say Online Reputation Management Is Important--Here's How To Keep Up*. https://www.forbes.com/sites/ryanerskine/2018/07/30/study-97-of-business-owners-say-online-reputation-management-is-important-heres-how-to-keep-up/?sh=35b06f1b6c02

Fairley, S. (2003). In search of relived social experience: Group-based nostalgia sport tourism. *Journal of Sport Management*. https://doi.org/10.1123/jsm.17.3.284

Garbe, W. (2019). *SymSpell: 1 Million times faster through symmetric delete spelling correction algorithm*. GitHub.

Ghasemi, S. Z., & Sanaei, M. (2015). Relationship between aesthetic sport facilities and customer satisfaction mediated by the quality of service in Mazandaran province. *International Journal of Sport Studies, 5*(7), 855–859.

Gibson, H. J. (1998). Sport tourism: A critical analysis of research. *Sport Management Review*. https://doi.org/10.1016/S1441-3523(98)70099-3

Glowacki, E. M., Glowacki, J. B., & Wilcox, G. B. (2018). A text-mining analysis of the public's reactions to the opioid crisis. *Substance Abuse, 39*(2), 129–133. https://doi.org/10.1080/08897077.2017.1356795

Goyal, P., Pandey, S., & Jain, K. (2018). SpaCy. In *Deep Learning for Natural Language Processing: Creating Neural Networks with Python*.

Gronroos, C. (1988). Service quality: The six criteria of good perceived service quality. *Review of Business, 9*(3), 10.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. *Tourism Management, 59*, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

Herrera, F., Herrera-Viedma, E., & Verdegay, J. L. (1995). A sequential selection process in group decision making with a linguistic assessment approach. *Information Sciences, 85*(4), 223–239. https://doi.org/10.1016/0020-0255(95)00025-K

Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2013.07.009

Sofotasiou, P., Hughes, B., & Calautit, J. K. (2014). Thermal performance evaluation of semi-outdoor sport stadia: A case study of the upcoming 2022 FIFA world cup in ... thermal performance evaluation of semi-outdoor sport Stadia: A case study of the upcoming 2022 FIFA world cup in Qatar. In *EinB2014*: *3 Rd International Conference "ENERGY in BUILDINGS 2014," 15*(October), 0–8.

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*.

Ibrahim, S. K. M., al Gharbawi, A. E. D., & Salam, E. A. (2020). The effect of the quality of services for sports facilities on the intention to use: Applying to sports facilities in the arab academy for science, technology and maritime transport—Egypt—Alexandria. *Oalib*. https://doi.org/10.4236/oalib.1106887

Iversen, E. B. (2015). Measuring sports facility utilisation by collecting performance information. *Managing Sport and Leisure*. https://doi.org/10.1080/23750472.2015.1090885

Kim, K., Park, O. J., Yun, S., & Yun, H. (2017). What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. *Technological Forecasting and Social Change*. https://doi.org/10.1016/j.techfore.2017.01.001

Kurtzman, J., & Zauhar, J. (2003). A wave in time: the sports tourism phenomena. *Journal of Sport and Tourism*. https://doi.org/10.1080/14775080306239

Lau, E., Hou, H. C., Lai, J. H. K., Edwards, D., & Chileshe, N. (2021). User-centric analytic approach to evaluate the performance of sports facilities: A study of swimming pools. *Journal of Building Engineering*. https://doi.org/10.1016/j.jobe.2021.102951

Lee, H., & Kim, C. (2014). Benchmarking of service quality with data envelopment analysis. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2013.12.008

Li, H., & Li, D. (2021). Recognition and optimization analysis of urban public sports facilities based on intelligent image processing. *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2021/8948248

Liang, D., Dai, Z., Wang, M., & Li, J. (2020). Web celebrity shop assessment and improvement based on online review with probabilistic linguistic term sets by using sentiment analysis and fuzzy cognitive map. *Fuzzy Optimization and Decision Making, 19*(4), 561–586. https://doi.org/10.1007/s10700-020-09327-8

Lin, M., Chen, Z., Xu, Z., Gou, X., & Herrera, F. (2021). Score function based on concentration degree for probabilistic linguistic term sets: An application to TOPSIS and VIKOR. *Information Sciences, 551*, 270–290. https://doi.org/10.1016/j.ins.2020.10.061

Liu, Y.-D., Taylor, P., & Shibli, S. (2009). Sport equity: Benchmarking the performance of english public sport facilities. *European Sport Management Quarterly*. https://doi.org/10.1080/16184740802461686

López Fernández, M. C., & Serrano Bedia, A. M. (2004). Is the hotel classification system a good indicator of hotel quality? An application in Spain. *Tourism Management*. https://doi.org/10.1016/j.tourman.2004.06.007

Losi, G., Bonzanini, A., Aquino, A., & Poesio, P. (2021). Analysis of thermal comfort in a football stadium designed for hot and humid climates by CFD. *Journal of Building Engineering*. https://doi.org/10.1016/j.jobe.2020.101599

Luca, M., & Reshef, O. (2021). The effect of price on firm reputation. *Management Science*. https://doi.org/10.1287/mnsc.2021.4049

Malik, M. I., Sindhu, M. A., Khattak, A. S., Abbasi, R. A., & Saleem, K. (2021). Automating test oracles from restricted natural language agile requirements. *Expert Systems*. https://doi.org/10.1111/exsy.12608

Mariani, M. M., & Visani, F. (2019). Embedding eWOM into efficiency DEA modelling: An application to the hospitality sector. *International Journal of Hospitality Management*. https://doi.org/10.1016/j.ijhm.2019.01.002

Masriki Frinaldi, A. (2021). The effect of service quality and facilities on visitor satisfaction at Gandoriah beach, Pariaman City. *International Journal of Tourism, Heritage and Recreation Sport, 2*(1), 9–14.

Moro, S., Lopes, R. J., Esmerado, J., & Botelho, M. (2020). Service quality in airport hotel chains through the lens of online reviewers. *Journal of Retailing and Consumer Services*. https://doi.org/10.1016/j.jretconser.2020.102193

Najafi, S., Saati, S., & Tavana, M. (2015). Data envelopment analysis in service quality evaluation: An empirical study. *Journal of Industrial Engineering International*. https://doi.org/10.1007/s40092-014-0090-6

Nie, R. X., Tian, Z. P., Kwai Sang, C., & Wang, J. Q. (2020). Implementing healthcare service quality enhancement using a cloud-support QFD model integrated with TODIM method and linguistic distribution assessments. *Journal of the Operational Research Society*. https://doi.org/10.1080/01605682.2020.1824554

Nilashi, M., Abumalloh, R. A., Alghamdi, A., Minaei-Bidgoli, B., Alsulami, A. A., Thanoon, M., Asadi, S., & Samad, S. (2021a). What is the impact of service quality on customers' satisfaction during COVID-19 outbreak? New findings from online reviews analysis. *Telematics and Informatics*. https://doi.org/10.1016/j.tele.2021.101693

Nilashi, M., Ahmadi, H., Arji, G., Alsalem, K. O., Samad, S., Ghabban, F., Alzahrani, A. O., Ahani, A., & Alarood, A. A. (2021b). Big social data and customer decision making in vegetarian restaurants: A combined machine learning method. *Journal of Retailing and Consumer Services*. https://doi.org/10.1016/j.jretconser.2021.102630

Niu, H., & Zhang, Y. (2021). The proportion of sports public service facilities based on the dea model in colleges and universities. *Revista Brasileira De Medicina Do Esporte*. https://doi.org/10.1590/1517-8692202127012020_0115

Núñez-Serrano, J. A., Turrión, J., & Velázquez, F. J. (2014). Are stars a good indicator of hotel quality? Assymetric information and regulatory heterogeneity in Spain. *Tourism Management*. https://doi.org/10.1016/j.tourman.2013.10.004

Oliveira, R., Pedro, M. I., & Marques, R. C. (2013). Efficiency and its determinants in Portuguese hotels in the Algarve. *Tourism Management*. https://doi.org/10.1016/j.tourman.2012.06.009

Oppermann, M. (2000). Tourism destination loyalty. *Journal of Travel Research*. https://doi.org/10.1177/004728750003900110

Park, J., & Lee, B. K. (2021). An opinion-driven decision-support framework for benchmarking hotel service. *Omega (united Kingdom), 103*, 102415. https://doi.org/10.1016/j.omega.2021.102415

Ramchandani, G., Shibli, S., & Kung, S. P. (2018). The performance of local authority sports facilities in England during a period of recession and austerity. *International Journal of Sport Policy*. https://doi.org/10.1080/19406940.2017.1420676

Ramchandani, G., & Taylor, P. (2011). Quality management awards and sports facilities' performance. *Local Government Studies*. https://doi.org/10.1080/03003930.2011.554824

Redmond, G. (1981). The world's sports museums. *Sports International*, 31–34.

Redmond, G. (1990). Points of increasing contact: Sport and tourism in the modern world. *Sport in Society: Policy, Politics and Culture*, *43*.

Santos, G., Marques, C. S., Justino, E., & Mendes, L. (2020). Understanding social responsibility's influence on service quality and student satisfaction in higher education. *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2020.120597

Sellers-Rubio, R., & Casado-Díaz, A. B. (2018). Analyzing hotel efficiency from a regional perspective: The role of environmental determinants. *International Journal of Hospitality Management*. https://doi.org/10.1016/j.ijhm.2018.03.015

Silva, C. G., & Ricketts, C. F. (2016). Mega sporting events and air pollution: The case of the super bowl. *Journal of Environmental Management and Tourism*. https://doi.org/10.14505/jemt.v7.3(15).16

Su, Y., & Teng, W. (2018). Contemplating museums' service failure: Extracting the service quality dimensions of museums from negative on-line reviews. *Tourism Management*. https://doi.org/10.1016/j.tourman.2018.06.020

Takata, K., & Hallmann, K. (2021). A systematic quantitative review of authenticity in sport tourism. *Journal of Sport and Tourism*. https://doi.org/10.1080/14775085.2021.1877564

Thu, H. N. T., Minh, T. T., Ngoc, T. N. T., Nguyen, B. G., & Ngoc, L. N. (2021). Measuring satisfaction and loyalty of guests based on vietnamese hotel online reviews. *International Journal of E-Entrepreneurship and Innovation*. https://doi.org/10.4018/IJEEI.2021070101

Tone, K., Chang, T. S., & Wu, C. H. (2020). Handling negative data in slacks-based measure data envelopment analysis models. *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2019.09.055

Vegara-Ferri, J. M., López-Gullón, J. M., Valantine, I., Suárez, A. D., & Angosto, S. (2020). Factors influencing the tourist's future intentions in small-scale sports events. *Sustainability (Switzerland)*. https://doi.org/10.3390/su12198103

Vimla, V., & Taneja, U. (2021). Brand image to loyalty through perceived service quality and patient satisfaction: A conceptual framework. *Health Services Management Research*. https://doi.org/10.1177/0951484820962303

Weed, M., & Bull, C. (2009). Sports tourism: Participants, policy and providers. *Sports Tourism: Participants Policy and Providers*. https://doi.org/10.4324/9780080474441

Wu, C.-C. (2011). The impact of hospital brand image on service quality, patient satisfaction and loyalty. *African Journal of Business Management*. https://doi.org/10.5897/AJBM10.1347

Wu, Q., Liu, X., Qin, J., Wang, W., & Zhou, L. (2021a). A linguistic distribution behavioral multi-criteria group decision making model integrating extended generalized TODIM and quantum decision theory. *Applied Soft Computing*. https://doi.org/10.1016/j.asoc.2020.106757

Wu, Q., Liu, X., Zhou, L., Tao, Z., & Qin, J. (2021b). A quantum framework for modeling interference effects in linguistic distribution multiple criteria group decision making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. https://doi.org/10.1109/TSMC.2021.3072388

Xiao, J., Wang, X., & Zhang, H. (2020). Managing personalized individual semantics and consensus in linguistic distribution large-scale group decision making. *Information Fusion*. https://doi.org/10.1016/j.inffus.2019.06.003

XLSTAT. (2007). *Statistical Software for Excel* (2021.3.1). Excel. https://www.xlstat.com

Xu, X., Yin, X., & Chen, X. (2019). A large-group emergency risk decision method based on data mining of public attribute preferences. *Knowledge-Based Systems*. https://doi.org/10.1016/j.knosys.2018.09.010

Yang, Z., Gao, Y., & Fu, X. (2021). A decision-making algorithm combining the aspect-based sentiment analysis and intuitionistic fuzzy-VIKOR for online hotel reservation. *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-04339-y

Yin, P., Chu, J., Wu, J., Ding, J., Yang, M., & Wang, Y. (2020). A DEA-based two-stage network approach for hotel performance analysis: An internal cooperation perspective. *Omega (United Kingdom)*. https://doi.org/10.1016/j.omega.2019.02.004

Yu, S. M., Wang, J., Wang, J. Q., & Li, L. (2018). A multi-criteria decision-making model for hotel selection with linguistic distribution assessments. *Applied Soft Computing Journal*. https://doi.org/10.1016/j.asoc.2017.08.009

Zhang, G., Dong, Y., & Xu, Y. (2014). Consistency and consensus measures for linguistic preference relations based on distribution assessments. *Information Fusion*. https://doi.org/10.1016/j.inffus.2012.01.006

Zhu, J. (2022). DEA under big data: Data enabled analytics and network data envelopment analysis. *Annals of Operations Research*. https://doi.org/10.1007/s10479-020-03668-8