J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 3823--3840

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on hotel customer satisfaction prediction and intelligent service management model based on deep neural network

Xiaolu Xu[1,✉]

[1] *School of Digital Tourism and Culture, Sichuan Vocational and Technical College, Suining, Sichuan, 629000, China*

ABSTRACT
With the rapid expansion of the Internet and e-commerce, and the rapid revolution of the consumption mode, customer reviews have become the most important feedback means for customers' preference and satisfaction level of products nowadays. In this paper, hotel customer reviews are used as the basis for predicting hotel customer satisfaction, and the TF-IDF feature word extraction method is proposed to extract review feature words. Based on deep neural network, we propose the sentiment analysis technology of hotel customer reviews, use BERT neural network to construct aspect term extraction model, realize the sentiment recognition and quantification of hotel customer reviews, and combine the fuzzy comprehensive evaluation and IPA analysis as the prediction and analysis model of hotel customer satisfaction. Taking 25837 customer reviews of XC Hotel as a research sample, we explore and analyze the satisfaction of XC Hotel customers. The secondary and primary features of the reviews were extracted by review feature words, and the themes were extracted by LDA theme mining model, which concluded that the evaluation items of concern for XC Hotel lie in location, facilities, hygiene, service, price, and food and beverage. The prediction results showed that 69.59% of customers were satisfied, 18.42% felt average and 11.99% were dissatisfied. IPA analysis of satisfaction and importance of XC Hotel and its visualization were conducted, and the intelligent service management model of the hotel was constructed based on the results of IPA analysis, and the optimization strategy of intelligent service of the hotel was proposed.

Keywords: TF-IDF, BERT, fuzzy comprehensive evaluation, IPA analysis, hotel customer satisfaction

1. Introduction

With the vigorous development of the national economy, the comprehensive national strength and the people's consumption level rising, the hotel industry in the people's life "food, clothing, housing and transportation" occupies a more important position. In the past, people pay more attention to the price

✉ Corresponding author.
E-mail address: luluchong1989@126.com (X. Xu).
Received 15 January 2024; Revised 10 February 2024; Accepted 30 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-217

and basic facilities of the hotel, however, with the popularization of the Internet and the improvement of people's living standards, the customer's needs and expectations of the hotel are also increasing. When choosing a hotel, customers not only pay attention to the cost-effectiveness of the hotel, but also to the hygiene environment, living comfort, service quality and other factors, they prefer to get a unique and pleasant experience in the hotel to meet their higher needs and expectations [1-3]. In order to meet the diversified needs of customers, hotel managers and practitioners need to continuously improve their own abilities and qualities, improve the quality of service management and optimize the level of product structure in order to enhance customer satisfaction [4-6].

Conducting hotel customer satisfaction and service management quality surveys is a key antecedent for hotels to cope with changes in the consumer market and consumption concepts [7-8]. Predicting hotel customer satisfaction provides a better understanding of customer preferences and behaviors, while intelligent service management can optimize the customer experience [9-10]. This information plays an important role in adjusting the hotel's marketing strategy, improving hardware facilities and service quality, and optimizing the room rate strategy, which can effectively enhance the hotel's customer satisfaction level [11-13]. Based on this, the proposed customer satisfaction enhancement strategy can not only reduce the operating costs and improve the profitability of the hotel, but also optimize the hotel staffing and innovate the hotel products and services, which is of great practical significance for the operation and management of the hotel [14-16].

The products and services provided by hotels are customer-experiential commodities, and the customer experience lasts for a long time, so predicting customer satisfaction using digital technology can help customers perceive the difference between the value of products and services and the expected value in advance. Literature [17] used online text review attributes and user community engagement to predict overall hotel customer satisfaction and based on this to develop effective online hotel management behaviors to improve the e-word of mouth and performance of hotels, the study showed that the subjectivity of the text reviews implies low satisfaction, while diversity and richness of emotions represent high satisfaction. Literature [18] combined deep learning techniques with expectation-confirmation theory to provide explanations for hotel customer perceptions using psychological and cognitive theories, and the study showed that the proposed predictive model for customer satisfaction based on customer reviews, hotel information and pictures has a high accuracy rate. Literature [19] shows that hotel electronic word-of-mouth is crucial in analyzing customer decision-making and enhancing hotel brand image, and the ANN-based hotel customer satisfaction prediction model is developed to identify important features related to customer satisfaction and provide theoretical suggestions for hotel service management. Literature [20] describes a method to improve hotel service management based on hotel customer experience and proposes a hotel customer satisfaction prediction model which uses a supervised machine learning approach to process a large amount of online review data and employs a plain algorithm to categorize the hotel reviews, and the statistical results of this model provide valuable insights for hotels to improve their services.

Hotels need to continuously optimize their service and management level with the help of intelligent technology to meet the diversified needs of customers, and then stand out in the fierce market competition to win more customers and market share. Literature [21] evaluated the performance of deep learning classification techniques based on convolutional neural network (CNN-DL) and support vector machine (SVM)-based classification techniques in making hotel service quality prediction, and analyzing hotel customers' comments with the help of digital technology and social media is conducive to advancing the hotel industry to improve customer experience. Literature [22] proposes a hotel service performance evaluation model, which first constructs a hierarchical ontology of hotel service defects, and then applies neural networks to identify the causes of hotel service defects in users' online reviews, so that hotel managers can understand customer needs and service assessment in real time. Literature [23] addresses the problem of hotel booking cancellation by building an artificial intelligence-based personal name record recognition model to strengthen the hotel booking management system by detecting individuals who cancel hotel service bookings within a short period of time, in order to reduce the losses of hotel companies. Literature [24] compares the attribute prediction performance of various deep learning classification algorithms, such as CNN-DL and SVM, in hotel evaluation, borrowing IoT technology to evaluate and classify the evaluation indexes of value, hygiene, and facilities of hotels, so that hotel enterprises can reduce the operation cost while maintaining customer satisfaction.

In this paper, hotel customer reviews are used as a medium to measure hotel customer satisfaction, and we propose word frequency statistics, TF-IDF, TextRank for feature word extraction of hotel customer reviews, and deep learning based on deep learning to realize sentiment analysis of hotel customer reviews. A joint model of text attribute extraction and sentiment recognition of hotel customer reviews based on BERT is established to extract the information attributes in the text and recognize the sentiment polarity in it, and identify the entity part and sentiment tendency in the text. Fuzzy comprehensive evaluation and IPA analysis are used as the main methods for hotel customer satisfaction evaluation prediction in this paper, and hotel customer satisfaction and importance are visualized by IPA with the help of the results of LDA topic mining model to calculate the weights. XC Hotel is selected as the research object, and its 25837 customer reviews are discussed and analyzed in depth for hotel customer satisfaction. Combined with the results of customer satisfaction prediction and analysis of XC Hotel, the intelligent service management model of the hotel is constructed, and the optimization strategy of intelligent service of the hotel is proposed in terms of customer interaction, application of technological ability of service personnel, and optimization of technological tools.

2. Hotel Customer Satisfaction Feature Extraction and Sentiment Analysis Methods

2.1. Hotel Customer Review Text Feature Word Extraction Technique

2.1.1. Definition of feature word extraction. Feature word extraction technique can be defined as a technical means to extract some special words or phrases from human language or speech that can represent the meaning of the whole sentence or are most relevant to the meaning of the whole sentence. Generally speaking, the common ways of feature word extraction are Word Frequency Statistics, TF-IDF, Text Rank.

2.1.2. Word frequency statistics. Word frequency statistics is the most traditional feature word extraction method, the idea is very simple, that is, it is considered that in a paragraph or an article, some important words tend to be mentioned repeatedly by the author in order to strengthen their importance. Word frequency-based feature word extraction assumes that these frequently occurring words can represent the meaning of the whole paragraph or article. Therefore, the frequency of occurrence of words in the article is counted and the N words with the highest frequency are selected to obtain the feature words. The formula for calculating the frequency is as follows:

$$f_p = \frac{n_w}{|N|} \tag{1}$$

where n_w is the number of occurrences of word w and $|N|$ is the total number of words.

2.1.3. Word frequency-inverse document frequency. Word Frequency-Inverse Document Frequency (abbreviated as TF-IDF) is used to calculate the importance score of a word in a sentence, paragraph or article [25].TF-IDF is based on the following assumption. Words with high criticality tend to be mentioned a lot in the text and words with low criticality tend to be less common in the text. At the same time, the more a word appears in a text, the less critical it is, and the less a word appears in a text, the more critical it is. If we want to measure the criticality of a word, we must consider both the frequency information of the word in the text and the frequency information of the word in the overall corpus. Therefore, TF-IDF is divided into two items, TF is called word frequency, which represents the ability of the word to summarize the sentence, and IDF is called inverse document frequency, which represents the ability of the word to distinguish the sentence. The calculation steps are as follows.

First, the frequency of the word's occurrence in the document is calculated:

$$tf_w = \frac{n_{dv}}{n_d} \tag{2}$$

where n_{dv} is the number of occurrences of word w in text d and n_d is the total number of words in text d. Then, the inverse document frequency of words in the corpus is calculated:

$$idf_w = \log\left(\frac{|D|}{|D_w|+1}\right) \tag{3}$$

where $|D|$ denotes the total number of documents in the corpus and $|D_w|$ denotes the number of documents in the corpus that contain the word w. Finally, the word frequency-inverse document frequency is computed:

$$tf - idf_w = tf_w \times idf_w \tag{4}$$

2.1.4. Text Rank. Text Rank is often used to extract keywords or key phrases. This method is evolved from the Page Rank algorithm, which is mainly modeled based on graph theory. The Page Rank algorithm considers both the number and the importance of links, and its core idea is that the more pages are linked to, the more important the page is, and the more important the page is linked to, the more important the page is [26]. Therefore, the web page ranking problem can be skillfully modeled as a graph model, where nodes represent web pages and edges represent the existence of links between web pages, i.e., if a web page links to other web pages, a directed edge from the node of that web page to the node of the linking web page is created. The node score formula is as follows:

$$S(v_i) = (1-t) + t * \sum_{j \in In(V_i)} \frac{1}{|Out(v_j)|} S(v_j) \tag{5}$$

where $S(v_i)$ is the importance score of web page i, t is the damping coefficient, $In(v_i)$ refers to the web page linked to web page i, and $Out(v_j)$ refers to the web page linked to web page j. The main reason for adding the damping factor is to ensure convergence, and the damping factor can be interpreted as the probability value that a user will go to a web page directly without going through a web page jump.

Compared with Page Rank, Text Rank also abstracts the problem of ranking words or phrases into a graph model, where a node represents a word or phrase, and an edge indicates a co-occurrence relationship between words or phrases, i.e., an edge exists between two words or phrases if the two words or phrases appear in the same window at the same time. The node score formula is as follows:

$$S(v_i) = (1-t) + t * \sum_{j \in In(V_i)} \frac{w_{ij}}{\sum_{k \in Out(V_j)} w_{jk}} S(v_j) \tag{6}$$

where $S(v_i)$ is the importance score of word or phrase i, t is the damping coefficient, $In(v_i)$ refers to the set of all words or phrases pointing to word or phrase i, $Out(v_j)$ refers to the set of other words or phrases pointing to word or phrase j, and w_{ij} denotes the weight of the edges between the two nodes, which can usually be determined by the number of co-occurrences.

2.2. Sentiment Analysis Techniques for Hotel Customer Reviews

2.2.1. Definition of Sentiment Analysis. The method of digging deeper into the human emotions involved in a text is known as sentiment analysis or emotional disposition analysis. There are three stages in the development of sentiment analysis techniques. The earliest was by summarizing the sentiment lexicon and making judgments about sentiment based on the lexicon rules. The next is to judge the sentiment by machine learning. At present, deep learning has become the mainstream trend.

2.2.2. Deep learning based approaches. Learning to make judgments about emotions. Currently, deep learning has become the mainstream trend. The use of deep learning for sentiment analysis has become mainstream and is mainly divided into two categories, convolutional neural networks, recurrent neural networks.

1) Convolutional Neural Network. As the name suggests, convolutional neural network (abbreviated as CNN) is a neural network with multi-layer convolutional operations in the forward propagation calculation process. CNNs are widely used in the era of deep learning because the convolution operation is very effective for data processing in the class of lattice structures [27]. The convolution and pooling operations of CNN networks alternate, and the number of features in the network usually tends to increase from left to right.

Convolution is the most important operation in CNN networks. The convolution kernel extracts features by continuously panning over the original mesh data and the discrete 2D convolution formula can be defined:

$$S(i,j)=\sum_{p}\sum_{q}K(p,q)*I(i-p,j-q) \tag{7}$$

where K denotes the convolution kernel, I denotes the matrix to be convolved, and S denotes the post-convolution matrix.

Convolution consists of three main features, translation isovariance, parameter sharing, and sparse links. Translation isomorphism means that if an object in the input is translated to another position, then the output features will be translated accordingly. Parameter sharing means that the same set of parameters is used for each operation as the convolution kernel is continuously translated. Sparse connectivity means that the number of input nodes linking each output node becomes less compared to a fully connected network. The role of the pooling layer is to perform feature dimensionality reduction and increase the model field of view. Due to the small number of parameters, the pooling operation can effectively control the complexity of the neural network and alleviate the overfitting phenomenon, maximum pooling and average pooling are the two most commonly used. The fully connected layer mainly plays the role of dimensionality reduction, which is convenient for the output layer to calculate the corresponding loss function.

2) Recurrent Neural Network. In many real-life situations, the structure of the data is not independent of each other but in a chain-like structure (there is mutual influence between neighboring features), and traditional fully connected networks are difficult to process this kind of data. In order to effectively process sequence-structured data, recurrent neural networks (abbreviated as RNN) have been proposed and widely used.

Input feature X_t at moment t and get the newly acquired information at the current moment after transformation by input layer weight matrix U. H_{t-1} is transformed by the hidden layer weight matrix W to get the previous moment's retained information. The new information of the current moment and the retained information of the previous moment are added together to get the hidden layer variable H_t of the current moment, and H_t is multiplied with the output layer weight matrix V to get O_t. The formula of RNN is as follows:

$$H_t=f\left(U\cdot X_t+W\cdot H_{t-1}\right) \tag{8}$$

$$O_t = g\left(V \cdot H_t\right) \tag{9}$$

2.3. Sentiment analysis method of hotel customer review information based on BERT

2.3.1. BERT-based Attribute Extraction and Sentiment Recognition. This subsection establishes a joint model for attribute extraction and sentiment recognition of hotel customer review text based on BERT for extracting the information attributes in the text and recognizing the sentiment polarity in it [28]. The system consists of two parts, the attribute term extraction model and the sentiment recognition model, the attribute term extraction model is responsible for identifying the attribute part of the review text, while the sentiment recognition model focuses on identifying the sentiment polarity of the attribute part. The goal of the system is to recognize both the entity part and the sentiment tendency in the text for a given comment text.

1)BERT-based attribute extraction. The aspect term extraction model is constructed using BERT neural network, aiming to identify and extract the attribute entity part in the review information of the hotel customer, in the order of attribute entity labeling, input vectorization, and attribute entity extraction model building, and the construction steps will be elaborated in detail next.

The first step is attribute entity labeling of review text. The attribute recognition model built in this section belongs to the training method of supervised learning, so the comment text dataset used for training is labeled with attributes to obtain a dataset that can be used for model training.

In the second step, input vectorization. Input vectorization converts text data into numeric vectors so that computers can process and analyze them, and the process includes the following steps. First, a character dictionary is constructed that maps each character to a random integer identifier. Then, each character in the text is converted to its corresponding identifier and the entire text is converted to a numeric vector. These vectors may have different lengths, so they need to be adjusted to the same length. For this reason, the researchers propose that vectors with lengths exceeding a certain threshold can be deleted and vectors with insufficient lengths can be expanded to the same length using padding methods. Word2Vec is then used to map all text vectors into vectors of lower dimensions of fixed length. This allows each text to be represented by a numeric matrix, which is then fed into BERT for word vector training for subsequent processing and analysis.

The third step is attribute entity extraction model construction. In this paper, we choose the BERT pre-trained language model to construct the comment text attribute entity extraction model, and the pre-trained BERT model is designed to improve the performance of most NLP tasks.In this paper, we deploy the BERT Encoder layer, which is designed to extract the contextual features of the comment information.For the BERT pre-trained model, the fine-tuning learning process is essential, and the BERT Encoder acts as the word embedding layer, the fine-tuning process is carried out independently according to the joint loss function of multi-task learning, X represents the tokenized input of the model, which can get the initial output of textual contextual features:

$$O_{BERT} = BERT(X) \tag{10}$$

where O_{BERT} is the output feature of the BERT model and BERT denotes the BERT Encoder layer.

In this paper, Linear and CRF are added to the fully connected layer, in which the Linear layer projects the input feature vectors into a higher-dimensional space to improve the expressiveness of the model, the CRF layer adds the constraint of label order to the sequence of results, and then inputs it into the attribute term extractor, and the activation function of the attribute term extractor is the softmax function, which first performs token-level classification for each token. Assuming that T_i represents the token at the corresponding location, then:

$$Y_{term} = \frac{\exp(T_i)}{\sum_{i=1}^{N} \exp(T_i)} \tag{11}$$

Where N denotes the number of token categories, Y_{term} denotes the token categories predicted by the attribute term extractor, and the text corresponding to the predicted attribute entities is subsequently extracted based on the probability values output by the activation function, the input layer is a vector representation of the converted words in the text comments, and these numeric vectors are input to the model in the order of the characters appearing in the comments, the BERT layer is essentially the The BERT layer is essentially a multidimensional hidden layer of the neural network, which accomplishes the updating of parameters and the learning of contextual features of the comment information.

2.3.2. Quantitative modeling of emotions. Through the emotion quantification model of single attribute of product and the attribute text clustering model, the emotion value of product sub-attributes can be calculated effectively, so as to realize the emotion judgment of product attributes and analyze the satisfaction of consumers on each attribute. However, to compare the user satisfaction of different products of the same type, it is necessary to construct the emotion quantification model of the product as a whole, and the emotion quantification from single attribute to the product as a whole is a recursive containment relationship, therefore, the overall emotion S_{p_m} of product p_m can be accurately quantified by Equation (12), where i is the total number of attributes:

$$S_{p_m} = \frac{\sum_{i=1}^{R} S_{p_m f_i}}{i} \tag{12}$$

3. Hotel Customer Satisfaction Evaluation Prediction and Analysis Model

3.1. *Fuzzy integrated evaluation method*

The fuzzy comprehensive evaluation method is based on the theory of fuzzy mathematics, which advocates the idea of using the affiliation function to describe the intermediate transition of the phenomenon differences [29]. According to the affiliation degree theory of fuzzy mathematics, the fuzzy comprehensive evaluation method transforms the evaluation of things from qualitative to quantitative, that is to say, the original more fuzzy factors, quantify them, and turn them into data that can be compared intuitively, i.e., defuzzification. In real life, most of the things or objects we study will be subject to a variety of factors, with fuzzy and difficult to quantify the difficulties, because fuzzy mathematics to study the fuzzy object is good at, so it has an unparalleled advantage when it comes to solving complex comprehensive evaluation problems. The park customer satisfaction studied in this paper belongs to the problem affected by a variety of factors, so the selection of fuzzy comprehensive evaluation method can achieve more satisfactory results. The method mainly includes the following six steps.

1) Determine the evaluation factor set of the research object. Factors are specific to the evaluation of the object of study, and are represented by U as the set of factors composed of various factors as set elements, $U = \{u_1, u_2, \cdots, u_m\}$. where u_i represents the i th factor affecting the object of study, which are more or less ambiguous.

2) Determine the value of the evaluation factors for the object of study. The specific evaluation factor value of the research object is the judgment made by the evaluator for each evaluation factor perceived, and the results of different judgments as elements constitute the evaluation factor value set, which is represented by V, $V = \{v_1, v_2, \cdots, v_n\}$. where V_j represents the j th evaluation result.

3) Carry out single-factor fuzzy evaluation to obtain the evaluation matrix. Using r_{i1} to denote the degree of affiliation of the i rd element in the factor set U to the 1st element in the factor value set V, the result of the single-factor evaluation of the 5th element can be expressed in terms of a fuzzy set $R_i = \{r_{i1}, r_{i2}, \cdots, r_{in}\}$. After obtaining the M single-factor evaluation set $\{R_1, R_2, \cdots, R_M\}$, it constitutes matrix R_{m*n}, which is called the fuzzy comprehensive evaluation matrix:

$$R = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_M \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} \tag{13}$$

4) Determine the weight vector of the evaluation factors. Since each factor has different degrees of importance in the mind of the evaluator, it is necessary to assign a specific weight a_i to u_i, and use A to represent the weight vector of each factor, i.e. $A = (a_1, a_2, \cdots, a_m)$. The weight size is determined and normalized, and the formula is $\sum^{m} a_i = 1$, $a_i >= 0$.

5) Determine the fuzzy comprehensive evaluation result vector. After determining the fuzzy weight vector A and obtaining the fuzzy transformed evaluation matrix R from U to V, the evaluation results of each factor are obtained by multiplying the two:

$$AR = (a_1, a_2, \cdots, a_m) \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{nn} \end{bmatrix} = (r_1, r_2, \cdots r_n) \tag{14}$$

6) De-fuzzy values, i.e., the fuzzy comprehensive evaluation set $B = AR$ and measurement scale H are used to calculate the comprehensive evaluation score E of the evaluation object:

$$E = B \times H \tag{15}$$

Where $H =$ (very dissatisfied, dissatisfied, generally satisfied, satisfied, very satisfied = (5, 4, 3, 2, 1)

7) Determine the system score and interpret the results.

3.2. IPA analysis method

IPA analysis, or Importance-Performance Analysis [30]. Its advantages are that it is intuitive, easy to use and easy to interpret. The result of IPA analysis is usually a four-quadrant matrix in which the individual evaluation variables are positioned according to their respective scores in quadrants with importance on the horizontal axis and satisfaction on the vertical axis.

In the IPA model plot, there exists the mean or median of each of the observed variables importance and satisfaction. A cross point will be formed in the graph with the value of the horizontal coordinate of the cross point being the mean (median) of importance and the value of the vertical coordinate being the mean (median) of satisfaction, and this point will divide the graph into four quadrants. The significance of the horizontal axis (I-axis) is the importance axis and the significance of the vertical axis (P-axis) is the satisfaction axis. Each of the observed variables is positioned one by one in the corresponding position in each of the four quadrants according to its importance and actual score of satisfaction.

In order to more reasonably reflect the role of each indicator on the evaluation object, it is necessary to determine the corresponding weight coefficients after selecting the evaluation indicators, and calculate the weights according to the results of the LDA theme mining model.The LDA theme mining results mean that each comment generates the text-topic probability value, and according to the magnitude of the probability value will be classified into the corresponding topic, and the probability value of each topic is not the same in different texts, therefore, the method firstly carries out denominator normalization. normalization, the specific steps are the first first make D is the set of all topics j, C is the set of all comments i, p_j^i is the LDA topic model to get the probability of comment i on topic j, and the second denominator normalization, its calculation formula is:

$$W = \sum_{j \in D} \sum_{i \in C} p_j^i \tag{16}$$

where D is the set of all topics j, C is the set of all comments i, p_j^i is the probability that the LDA topic model obtains comments i about topic j, and finally the weights obtained for each topic W_j are calculated in Eq:

$$W_j = \frac{\sum_{i \in C} p_j^i}{W} \tag{17}$$

4. XC Hotel Customer Satisfaction Prediction Analysis

XC Hotels is an influential high-end art hotel brand in the Chinese market.XC Hotels initially opened only a few hotels in major cities such as Beijing and Shanghai, but as the market demand grew, the brand began to accelerate the pace of expansion, entering China's second-tier cities and international markets in 2015 and 2018. To date, XC Hotels has opened more than 1,000 hotels worldwide and is still expanding.

By continuously improving customer satisfaction, the room occupancy rate can be increased, thus improving the overall efficiency of the hotel. In order to achieve this goal, this paper will explore and analyze the customer satisfaction situation of XC Hotel based on 25,837 customer reviews of XC Hotel, which will provide a data basis for the subsequent proposal of the intelligent service management model of the hotel.

4.1. Hotel Customer Review Feature Extraction

4.1.1. Secondary comment feature extraction. In this section, feature words of XC Hotel customer reviews are extracted using the TF-IDF feature word extraction method proposed in this paper, and TF-IDF values are computed for each word after text preprocessing. First, the existing corpus data is used for training. Then, the TF-IDF algorithm is loaded to train the corpus data of this paper and only nouns and gerunds are retained as keywords by means of lexical filtering. Finally, the words are arranged in descending order according to the TF-IDF values, and the top 50 words are extracted from the vocabulary and used as the secondary features of XC hotel customer reviews, the details of which are shown in Table 1.The top ten feature words in terms of TF-IDF values are clean, breakfast, front desk, enthusiasm, location, facilities, satisfaction, neatness, service attitude, and transportation, respectively.

Table 1. The secondary characteristics of hotel customer reviews

Number	Key words	TF-IDF	Number	Key words	TF-IDF
1	Clean	0.355438	26	Dining room	0.023838
2	Breakfast	0.25407	27	Lobby	0.023562
3	Foreground	0.218278	28	Gym	0.022238
4	Enthusiasm	0.119977	29	Species	0.022019
5	Position	0.112338	30	Warmth	0.021081
6	Facilities	0.098022	31	Sound insulation	0.020531
7	Satisfaction	0.080223	32	Outbreak	0.018927

8	Neatness	0.079248	33	Air conditioning	0.016798
9	Service attitude	0.076431	34	Pillow	0.015709
10	Traffic	0.072885	35	Subway outlet	0.015097
11	Intimate	0.06767	36	Mask	0.014976
12	Quietness	0.053535	37	Mineral water	0.014949
13	Parking lot	0.045593	38	Style	0.01487
14	Performance ratio	0.045156	39	Robots	0.013692
15	Like	0.044482	40	Elevator	0.012921
16	Convenience	0.044045	41	Toilet	0.011466
17	Worthy of	0.039376	42	Mattress	0.011497
18	Steward	0.033079	43	Hot water	0.011479
19	Taste	0.026807	44	Television	0.010683
20	Scenic spot	0.025678	45	Fruit	0.010384
21	Decorate	0.024978	46	Alcohol	0.009573
22	Perimeter	0.024671	47	Shopping center	0.009353
23	Laundry room	0.024587	48	Odour	0.009226
24	Price	0.024115	49	Carefulness	0.008998
25	Good service	0.024003	50	Network	0.008344

4.1.2. First-level comment feature extraction. K-means clustering algorithm has the advantages of simplicity, applicability to different types of data, handling large amount of data, fast convergence, etc. k-means clustering algorithm is suitable for dealing with uniformly distributed datasets and can handle high dimensional data. In this study, K-means clustering algorithm is used to calculate the Euclidean distance between word vectors and to obtain the clustering results of word vectors. Contour coefficients are used to assess the quality of clustering. Repeated experiments are conducted by adjusting the K value, and it is found that the contour coefficient reaches a maximum value of 0.673 when the K value is 8, which indicates that the K value is optimal at this time, and the clustering results are shown in Fig. 1.

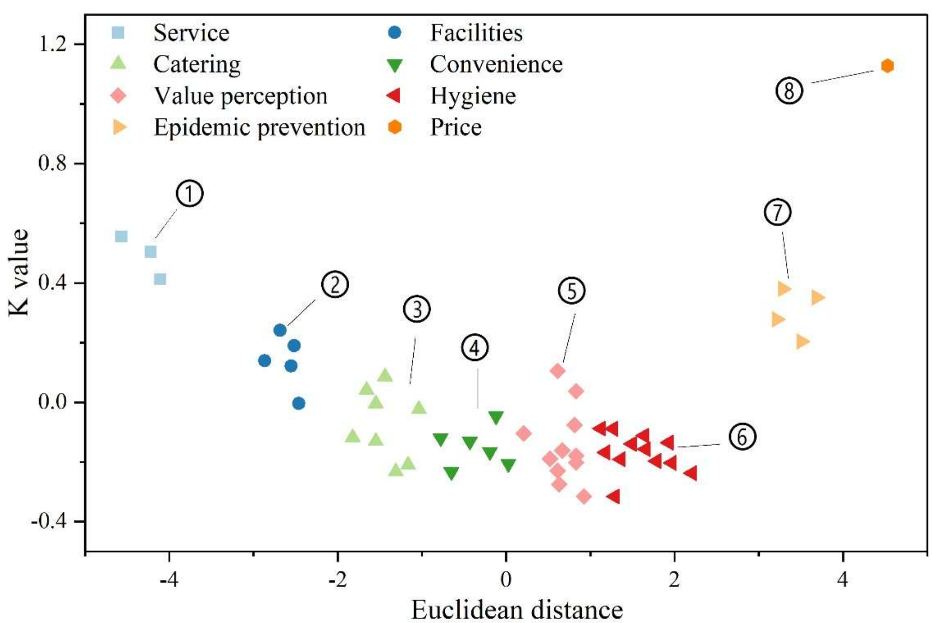

Fig. 1. Cluster result

In the above clustering, data points near the same center of mass are assigned to the same category. Based on the clustering results of the second-level features, we divided the first-level features of hotel customer reviews into eight categories, which are service, facilities, catering, convenience, value perception, hygiene, epidemic prevention and price, and the specific classification results are shown in Table 2.

Table 2. Classification result

Primary characteristic	Secondary characteristics
Service	Front desk, service attitude, waiter, service, lobby, robot
Facilities	Facilities, laundry rooms, gym, soundproof, air conditioning, pillows, elevators, bathrooms, mattresses, hot water, television, network
Catering	Breakfast, restaurant, type, mineral water, fruit
Convenience	Location, transportation, parking lot, convenience, scenic spots, surrounding, subway entrance, shopping center
Value perception	Warm, sweet, quiet, cost-effective, worthy, decorated, warm, style, and careful
Hygiene	Clean, tidy, smelly, smelly
Epidemic prevention	Outbreak, mask, alcohol
Price	Price

4.2. *Hotel Review LDA Topic Extraction*

Based on the LDA theme mining model, the theme extraction of XC Hotel review words was performed, and the resulting theme categories and their corresponding theme words are shown in Table 3 (the number after the theme word indicates the probability that the theme word belongs to a specific theme).The themes extracted by LDA are the attributes of this tourism product that customers are concerned about in their subjective writing in the reviews. The results show that the evaluation items that customers are concerned about XC Hotel are location, facilities, hygiene, service, price and food and beverage.

Table 3. LDA topic extraction

TOPIC1		TOPIC2		TOPIC3		TOPIC4		TOPIC5		TOPIC6	
Geographical location		Facilities		Hygiene		Service		Price		Catering	
Traffic	0. 066	Air conditioni ng	0. 042	Taste	0. 029	Service attitude	0. 044	Worthy	0. 024	Breakf ast	0. 022
IFS	0. 044	Facilitie s	0. 03	Clean	0. 025	Foregro und	0. 029	Price	0. 019	Dining room	0. 021
Position	0. 036	Networ k	0. 021	Conve nience	0. 021	Attitud e	0. 021	Height	0. 015	Cake	0. 017
Conveni ence	0. 023	Gym	0. 02	Not bad	0. 019	Nation al gold	0. 02	Performan ce ratio	0. 015	Give away	0. 017
Conveni ence	0. 22	Laundry room	0. 02	Stewa rd	0. 015	Stewar d	0. 018	Preferentia l treatment	0. 015	Fruit	0. 017
Downto	0.	Toilet	0.	Swee	0.	Check-	0.	IFS	0.	Recom	0.

wn	017		02	p	011	out	017		013	mend	016
Subway	0. 011	Hot water	0. 013	Neatn ess	0. 011	Conven ience	0. 015	Not bad	0. 012	Conve nience	0. 013
Center	0. 007	Laundry room	0. 011	Room	0. 01	Robots	0. 011	Position	0. 004	Miner al water	0. 009
Parking lot	0. 009	Facilitie s	0. 01	Cente r	0. 008	Lobby	0. 006	Center	0. 004	Servic e	0. 006
Shoppi ng center	0. 003	Sound insulation	0. 006	Servi ce	0. 005	Clean	0. 006	Service	0. 013	Worth y	0. 005

4.3. Customer satisfaction measurement based on fuzzy comprehensive evaluation

In order to measure customer satisfaction based on fuzzy comprehensive evaluation, it is necessary to analyze the XC Hotel reviews beforehand by using text analysis techniques and integrating them to produce a table of fuzzy evaluation based data for XC Hotel reviews, as shown in Table 4. The table analyzes the positive, neutral, and negative reviews to make factor judgments on the evaluation items, and the result is that 69.59% of the customers expressed satisfaction, 18.42% of the customers felt average, and 11.99% of the customers were dissatisfied. In this study, the customer satisfaction of XC Hotel was determined to be "satisfied" according to the principle of "maximum affiliation", based on the principle of "maximum affiliation", based on the fuzzy comprehensive evaluation set of positive, neutral and negative emotions assigned to 5, 3 and 1 respectively, and the six factors of geographical location, facilities, hygiene, service, price and food and beverage were integrated. The overall rating of the six evaluation items, namely, location, facilities, hygiene, service, price and catering, was also “satisfactory”.

Table 4. Fuzzy comprehensive evaluation

Evaluation project	Evaluation factor	General evaluation clause	Positive emotional comment clause	Neutral emotional comment clause	Negative emotional comment clause
			69.59%	18.42%	11.99%
Geographic location	Traffic	300	280	17	3
	IFS	190	168	20	2
	Position	749	496	223	30
	Convenience	181	121	52	8
	Convenience	253	250	2	1
	Downtown	2178	1889	274	15
	Subway	157	122	34	1
	Center	2896	2692	203	1
The total word frequency and proportion of geographical position	-	6904	6018	825	61
	-	26.72%	33.47%	17.33%	1.97%
Facilities	Air conditioning	4820	3142	1038	640
	Facilities	101	23	58	20
	Network	1178	896	202	80
	Gym	124	34	24	66
	Laundry room	143	38	5	100
	Toilet	73	15	9	49
	Hot water	139	11	3	125
	Laundry room	59	1	27	31
The total word frequency and proportion of facilities	-	6637	4160	1366	1111
	-	25.69%	23.14%	28.7%	35.87%
Hygiene	Taste	200	110	24	66
	Clean	645	591	23	31
	Convenience	177	161	7	9
	Not bad	169	65	24	80
	Steward	167	44	70	53
The total word frequency and proportion of health	-	1358	971	148	239
	-	5.26%	5.4%	3.11%	7.72%
Service	Service attitude	78	60	3	15
	Foregro	54	14	29	11

	und				
	Attitude	449	87	144	218
	National gold	3875	2651	953	271
	Steward	307	100	146	61
	Check-out	187	153	25	9
	Convenience	1186	897	175	114
The total word frequency and proportion of service	-	6136	3962	1475	699
	-	23.75%	23.04%	30.99%	22.57%
Price	Worthy	297	206	21	70
	Price	13	4	1	8
	Height	24	1	14	9
	Performance ratio	274	140	62	72
	Preferential treatment	216	111	57	48
The total word frequency and proportion of price	-	824	462	155	207
	-	3.19%	2.57%	3.26%	6.68%
Catering	Breakfast	1574	688	446	440
	Dining room	271	151	62	58
	Cake	662	472	93	97
	Give away	1340	989	179	172
	Fruit	131	108	10	13
The total word frequency and proportion of catering	-	3978	2408	790	780
	-	15.4%	13.39%	16.6%	25.19%

4.4. Hotel Customer Satisfaction IPA Analysis

This paper draws on IPA analysis to analyze the satisfaction levels of different hotel customer review characteristic items. In order to make the comparison of the satisfaction level of each item index more intuitive, SPSS statistical analysis software is used to conduct IPA analysis and its visualization of the satisfaction and importance of XC Hotel. The horizontal axis is the degree of importance and the vertical axis is the degree of actual satisfaction performance. The matrix is divided into four quadrants, and the importance and satisfaction values of the six evaluation items, namely, location, facilities, hygiene, service, price, and catering, are respectively projected into the IPA matrix, and the IPA distribution chart formed is shown in Figure 2.

1) IPA analysis method of the first quadrant (advantageous area) analysis. The first quadrant is the dominant region in the IPA analysis, i.e., the quadrant with high importance and also high scores on the actual performance of customer satisfaction. This quadrant indicates that customers pay significantly more attention to the factors of location and service. For XC Hotel, except for the objective attribute of "geographic location", which is difficult to change, XC Hotel needs to continue to pay more attention to the service factor and keep maintaining and improving the service of the hotel.

2) IPA analysis method of the second quadrant (repair area or improvement area) analysis. The second quadrant (focus on improvement area) This area represents the horizontal coordinate importance index is very high, at the same time the actual customer satisfaction score is low. Then it also indicates that the factors falling in this quadrant are particularly important for improving the hotel's intelligent services. In the eyes of the customer these important factors are not being met. As can be seen from the graph, "Facilities" has a high level of importance but a low level of customer satisfaction. XC hotel managers should focus on strengthening the construction of infrastructure improvement, sound infrastructure is to meet the material needs of hotel customers to ensure.

3) IPA analysis method of the third quadrant (opportunity zone) analysis. The third quadrant is also known as the opportunity zone, this quadrant reflects that customers believe that the importance of the region in the minds of consumers is not high, while customers are not optimistic about the actual satisfaction. Although the importance of this area is not as high as the first and second quadrants, but if we can improve the actual performance score of this area will also help to improve the overall XC hotel intelligent hotel will have new inspiration. XC hotel managers can improve the factors in the third quadrant according to their actual ability, improve the quality of "food and beverage" and establish a reasonable "price" system.

4) IPA analysis of the fourth quadrant (retention area) analysis. The holding area in the fourth quadrant indicates that customers think that the factors falling in this area do not have high importance, but after experiencing it, they feel that the actual performance of the satisfaction is very high, this part of the hotel intelligent service can not provide immediate improvement strategies but can give the hotel a hint of what the hotel has done well in the intelligent service. XC Hotel has been widely praised by customers for its performance in "hygiene", and it should continue to maintain its hygienic and clean characteristics as a long-term competitive point for the hotel.

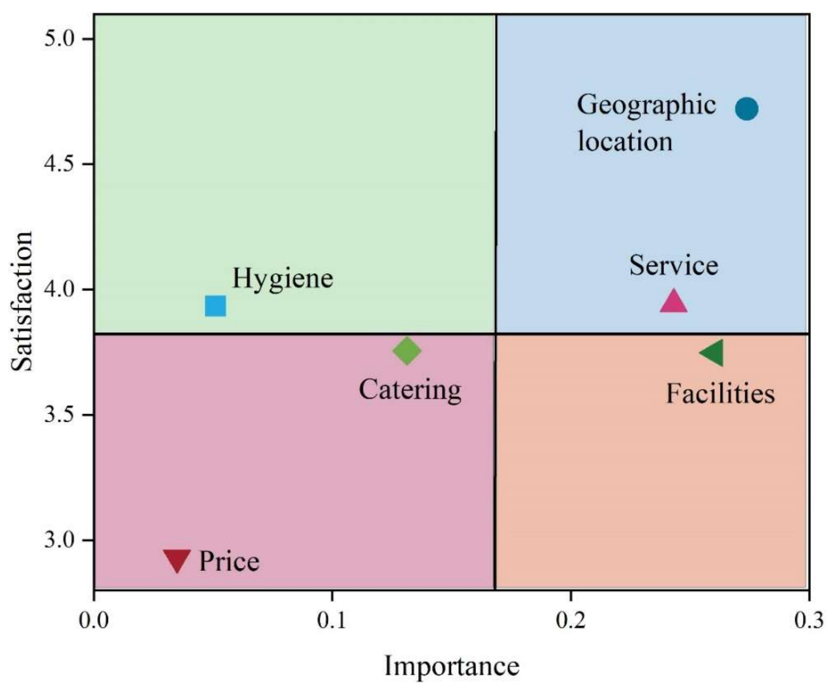

Fig. 2. IPA distribution

5. Intelligent Service Management Model for Hotels Based on Customer Satisfaction

Based on the previous IPA analysis of XC Hotel's customer satisfaction, this chapter will construct its hotel intelligent service management model and propose optimization strategies for hotel intelligent service from the perspective of hotel intelligent development.

1) Optimization strategy of customer interaction. Strengthen the promotion of customer interaction platform and formulate the customer interaction return visit model. Through the official website of Hotel H, WeChat, microblogging and other Internet channels to provide visual advertising interface to introduce the implementation of the technical tools and coverage and coverage of store information. In conjunction with the hotel's internal wall advertisements or electronic screens to maximize the conversion rate of the promotion. Customer interaction model is divided into two parts, the first part has been mentioned in the interface promotion, combined with the interaction of major online platforms, there is a centralized processing station to deal with and respond to online platforms and consulting in the message of the customer to stay in the problem and suggestions. The other part is the mode of regular interaction interface within the hotel, in the recent new installation of the whole-house intelligent voice control system in the H hotel, the hotel arranged the experience of the sense of the message area, to collect this part of the first experience of the intelligent service of the occupants of the customer's opinions.

2) Optimization strategy for the application of technical ability of service personnel. Unify the hotel internal training standards, standardize the hotel internal service level. In the hotel service personnel to get standardized on the basis of consistent training, and further will have the ability to optimize the training content of the management staff, so that the part of the staff in addition to mastering the operation of the technical tools of the process also need to have when the technical tools need to be equipped with the service mode.

3) Optimization of technical tools. Realize the stability of technical tools and the simple operation of technical tools. In the process of balancing the input cost and output of the hotel, it is also worthwhile for Hotel H to maintain and optimize the stability of the technical system by providing high-quality technical services to the customers. The ease of use and the ease of experience that can be realized through the use of these tools make it possible for the guests to immediately remember the advantages of this technological innovation. For customers of different ages, the technical performance is really easy to operate.

4) Optimization Strategy of Intelligent Service Content. Maintain the uniqueness of intelligent services, form the brand culture of intelligent services, and carry out continuous intelligent service content enhancement. The hotel can transform the investment in the uniqueness of intelligent services into customer customization, combine the concept of intelligent services with the hotel's own development, seek common ground while reserving differences, and at the same time continuously improve the content of intelligent services, so as to become a new image independently defined by the customers staying in the hotel.

6. Conclusion

This paper takes hotel customer reviews as the basis of hotel customer satisfaction prediction, extracts hotel customer review text feature words using Text Rank, Text Rank and other techniques, and based on this, combines deep neural network to mention BERT-based hotel customer review information sentiment analysis method, and formulates hotel customer satisfaction evaluation prediction method mainly based on fuzzy comprehensive evaluation and IPA analysis. XC Hotel is taken as the research object of hotel customer satisfaction prediction in this study, and its 25837 customer reviews are taken as the research samples to carry out in-depth hotel customer satisfaction prediction analysis. The feature words of XC Hotel's customer reviews are extracted, and the top ten secondary feature words of TF-IDF value are clean, breakfast, front desk, enthusiasm, location, facilities, satisfaction, neatness, service attitude, transportation, and the first-level features of eight categories are obtained with the help of K-means clustering algorithm: service, facilities, food and beverage, convenience, value perception, hygiene, epidemic prevention, and price. Further, subject extraction of review features is performed by LDA subject mining model to obtain 6 evaluation items of location, facilities, hygiene, service, price and catering. The evaluation items were judged by fuzzy comprehensive evaluation analysis, and the customers who expressed satisfaction, average and dissatisfaction accounted for 69.59%, 18.42% and 11.99%, respectively. IPA visualization analysis was carried out using IPA analysis for six evaluation items: geographic location, facilities, hygiene, service, price and catering. Geographic location and service are located in the first quadrant (dominance zone), food and beverage and price are both in the third quadrant (opportunity zone), while facilities and hygiene are located in the second quadrant (repair or improvement zone) and the fourth quadrant (maintenance zone), respectively. Drawing on the results of this paper's analysis of XC Hotel's customer satisfaction prediction, we constructed a hotel intelligent service management model based on customer satisfaction, and proposed a hotel intelligent service optimization strategy centered on customer interaction, service personnel's technological capabilities, technological tools, and intelligent service content.

References

[1] Aakash, A., & Gupta Aggarwal, A. (2022). Assessment of hotel performance and guest satisfaction through eWOM: big data for better insights. International Journal of Hospitality & Tourism Administration, 23(2), 317-346.

[2] Radojevic, T., Stanisic, N., & Stanic, N. (2017). Inside the rating scores: A multilevel analysis of the factors influencing customer satisfaction in the hotel industry. Cornell Hospitality Quarterly, 58(2), 134-164.

[3] Bi, J. W., Liu, Y., Fan, Z. P., & Zhang, J. (2020). Exploring asymmetric effects of attribute performance on customer satisfaction in the hotel industry. Tourism Management, 77, 104006.

[4] El-Adly, M. I. (2019). Modelling the relationship between hotel perceived value, customer satisfaction, and customer loyalty. Journal of Retailing and Consumer Services, 50, 322-332.

[5] Mehta, M. P., Kumar, G., & Ramkumar, M. (2023). Customer expectations in the hotel industry during the COVID-19 pandemic: a global perspective using sentiment analysis. Tourism Recreation Research, 48(1), 110-127.

[6] Kim, Y. J., & Kim, H. S. (2022). The impact of hotel customer experience on customer satisfaction through online reviews. Sustainability, 14(2), 848.

[7] Apriliani, D., Abidin, T., Sutanta, E., Hamzah, A., & Somantri, O. (2021). SentiHotel: a sentiment analysis application of hotel services using an optimized neural network. Bulletin of Electrical Engineering and Informatics, 10(3), 1536-1545.

[8] Wang, J., Zhao, Z., Liu, Y., & Guo, Y. (2021). Research on the role of influencing factors on hotel customer satisfaction based on BP neural network and text mining. Information, 12(3), 99.

[9] Bian, Y., Ye, R., Zhang, J., & Yan, X. (2022). Customer preference identification from hotel online reviews: A neural network based fine-grained sentiment analysis. Computers & Industrial Engineering, 172, 108648.

[10] Guerra-Montenegro, J., Sanchez-Medina, J., Laña, I., Sanchez-Rodriguez, D., Alonso-Gonzalez, I., & Del Ser, J. (2021). Computational Intelligence in the hospitality industry: A systematic literature review and a prospect of challenges. Applied Soft Computing, 102, 107082.

[11] Knani, M., Echchakoui, S., & Ladhari, R. (2022). Artificial intelligence in tourism and hospitality: Bibliometric analysis and research agenda. International Journal of Hospitality Management, 107, 103317.

[12] Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y., & Gupta, B. (2018). Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. Journal of computational science, 27, 386-393.

[13] Muhammad, P. F., Kusumaningrum, R., & Wibowo, A. (2021). Sentiment analysis using Word2vec and long short-term memory (LSTM) for Indonesian hotel reviews. Procedia Computer Science, 179, 728-735.

[14] Lukanova, G., & Ilieva, G. (2019). Robots, artificial intelligence, and service automation in hotels. In Robots, artificial intelligence, and service automation in travel, tourism and hospitality (pp. 157-183). Emerald Publishing Limited.

[15] Zhong, L., Zhang, X., Rong, J., Chan, H. K., Xiao, J., & Kong, H. (2021). Construction and empirical research on acceptance model of service robots applied in hotel industry. Industrial Management & Data Systems, 121(6), 1325-1352.

[16] Piccoli, G., Lui, T. W., & Grün, B. (2017). The impact of IT-enabled customer service systems on service personalization, customer service perceptions, and hotel performance. Tourism Management, 59, 349-362.

[17] Zhao, Y., Xu, X., & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. International journal of hospitality management, 76, 111-121.

[18] Oh, S., Ji, H., Kim, J., Park, E., & del Pobil, A. P. (2022). Deep learning model based on expectation-confirmation theory to predict customer satisfaction in hospitality service. Information Technology & Tourism, 24(1), 109-126.

[19] Aakash, A., Tandon, A., & Gupta Aggarwal, A. (2021). How features embedded in eWOM predict hotel guest satisfaction: An application of artificial neural networks. Journal of Hospitality Marketing & Management, 30(4), 486-507.

[20] Sánchez-Franco, M. J., Navarro-García, A., & Rondán-Cataluña, F. J. (2019). A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services. Journal of Business Research, 101, 499-506.

[21] Rajesh, S., Abd Algani, Y. M., Al Ansari, M. S., Balachander, B., Raj, R., Muda, I., ... & Balaji, S. (2022). Detection of features from the internet of things customer attitudes in the hotel industry using a deep neural network model. Measurement: Sensors, 22, 100384.

[22] Chiang, T. A., Che, Z. H., Huang, Y. L., & Tsai, C. Y. (2022). Using an ontology-based neural network and DEA to discover deficiencies of hotel services. International Journal on Semantic Web and Information Systems (IJSWIS), 18(1), 1-19.

[23] Sánchez, E. C., Sánchez-Medina, A. J., & Pellejero, M. (2020). Identifying critical hotel cancellations using artificial intelligence. Tourism Management Perspectives, 35, 100718.

[24] Shoukry, A., & Aldeek, F. (2020). Attributes prediction from IoT consumer reviews in the hotel sectors using conventional neural network: deep learning techniques. Electronic Commerce Research, 20(2), 223-240.

[25] C.A Nurhaliza Agustina,Rice Novita,Mustakim & Nesdi Evrilyan Rozanda. (2024). The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm. Procedia Computer Science156-163.

[26] M. A. H. Wadud,M. A. Jafor,M. F. Mridha & M. M. Rahman. (2020). Similarity Measurement Technique for Measuring the Performance of Page Rank Algorithm Based on Hadoop. International Journal of Recent Technology and Engineering (IJRTE)(5),4712-4717.

[27] Bhuvaneshwari P.,Rao A. Nagaraja,Robinson Y. Harold & Thippeswamy M. N. (2022). Sentiment analysis

for user reviews using Bi-LSTM self-attention based CNN model. Multimedia Tools and Applications(9),12405-12419.

[28] Nossayba Darraz,Ikram Karabila,Anas El Ansari,Nabil Alami & Mostafa El Mallahi. (2025). Integrated sentiment analysis with BERT for enhanced hybrid recommendation systems.Expert Systems With Applications125533-125533.

[29] Anqi Guo,Peiling He,Qing Min & Shuqiang Xu. (2024). A fuzzy comprehensive evaluation method for assessing the damage degree of cultural relics based on an improved AHP-EWM coupled weight model and a modified ridge function. Heritage Science(1),378-378.

[30] Lai Qi & Denholm John. (2024). Importance-Performance Analysis (IPA) in Researching the Satisfaction of Simulation. International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)(1),1-18.