Advances in Technology Innovation, vol. 10, no. 3, 2025, pp. 309-320

Effective Recommendation Considering Customers' Needs Using Review Texts with TF-IDF and Word2Vec: Case of Golf Course

Kodai Kimura[1,*], Takashi Namatame[2]

[1]Graduate School of Science and Engineering, Chuo University, Tokyo, Japan

[2]Faculty of Science and Engineering, Chuo University, Tokyo, Japan

Received 30 December 2024; received in revised form 16 May 2025; accepted 20 May 2025

DOI: https://doi.org/10.46604/aiti.2024.14699

Abstract

This paper aims to recommend the most suitable golf course for each user by focusing on golf courses and analyzing customer reviews. Furthermore, by examining the recommendation results, the goal is to clarify the characteristics of each golf course from the user's perspective and contribute to the promotion of each golf course. The procedure of this paper is first to extract user preferences using Word2vec and TF-IDF from reviews. Next, the extracted user preferences are matched with golf course features. Finally, recommendations are made based on the geographical relationship between the user and the golf course. As a result, a high accuracy rate is achieved. Additionally, some keywords that should be used in promotions for each golf course feature have been identified.

Keywords: recommendation, golf course, Word2Vec, TF-IDF

1. Introduction

User reviews have rapidly spread with the rise of the Internet. Especially on social media and e-commerce platforms, consumers can easily share opinions and experiences, making reviews a key source of information for purchasing decisions. According to the Ministry of Internal Affairs and Communications [1], over two-thirds of users across all age groups rely on reviews when shopping online (Fig. 1), highlighting their strong influence on consumer behavior. Accordingly, reviews are often utilized in recommendation systems for their ability to reflect consumer preferences.

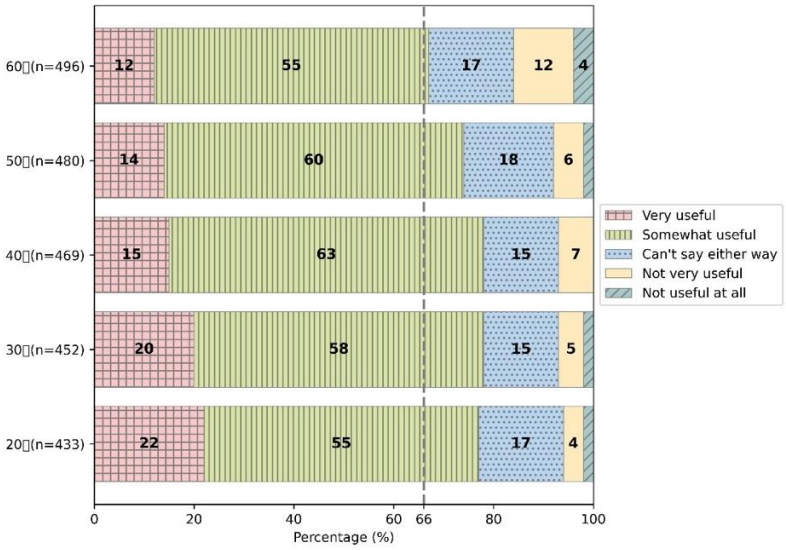

Fig. 1 Survey results: Do you refer to reviews when shopping online?

* Corresponding author. E-mail address: nama@kc.chuo-u.ac.jp

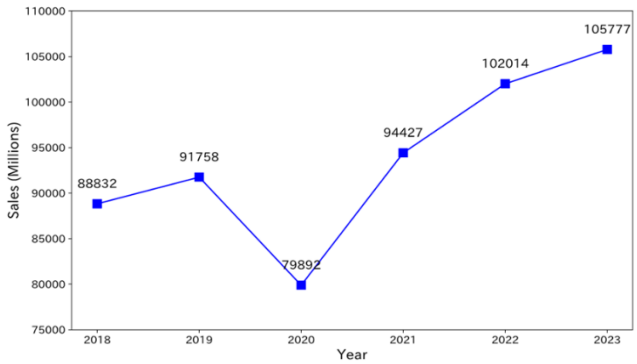

Fig. 2 Revenue of Golf Courses

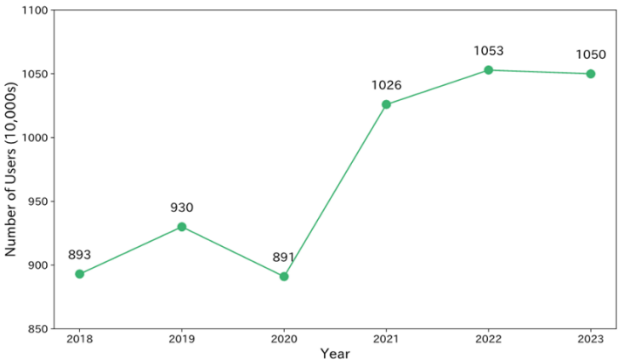

Fig. 3 Number of Golf Course Users

The Ministry of Economy, Trade and Industry [2] reported that Japan's golf industry experienced a sharp decline in 2020 due to COVID-19 but recovered by 2023 to levels exceeding those of 2019 (Figs. 2 and 3). This suggests a positive outlook for the industry.

Various studies have explored golf-related topics. Bekken et al. [3] evaluated the eco-efficiency of golf courses using profit and play volume. Izumi et al. [4] studied how golf helps female CEOs build male-dominated networks. Edwards et al. [5] examined the link between swing techniques and back pain. For recommendation systems, Khazaeli et al. [6] proposed an AI-based club selection tool, while Morise et al. [7] investigated collaborative filtering with deep learning using multi-criteria evaluation data. However, few studies have focused on golf course recommendations using reviews.

In other domains, reviews have been widely used in recommendations. Asani et al. [8] extracted food preferences from reviews for restaurant suggestions. Elahi et al. [9] integrated sentiment analysis into hybrid recommendation models. Abbasi-Moud et al. [10] developed a tourism recommender using user preferences. Syafganti et al. [11] found that review ratings and reviewer expertise affect hotel booking intentions. Wang et al. [12] showed that emotional factors in reviews—such as pleasure, trust, anger, and disgust—have an impact on user intentions.

To leverage reviews, extracting user preferences is key. Common methods include TF-IDF and Word2Vec. Zhou et al. [13] compared these methods in response prediction tasks. Zhang et al. [14] applied Word2Vec for medical term extraction. Al Tawil et al. [15] used TF-IDF, Word2Vec, and BERT to detect phishing. Liyih et al. [16] conducted sentiment analysis on YouTube comments using word embeddings. Abid et al. [17] extracted music-related keywords from Twitter using TF-IDF and Loglikelihood.

To fill this research gap, this study uses reviews to recommend golf courses by applying TF-IDF and Word2Vec to extract user preferences. Analyzing golf course reviews in this way is expected to generate recommendations aligned with user needs.

This study focuses on reviews posted by users who have made reservations and played at golf courses. Using these reviews, it aims to extract user preferences and recommend suitable golf courses by matching the extracted features with users' needs. Additionally, based on the analysis, this study seeks to identify effective promotional keywords for each golf course, contributing to the improvement of marketing strategies in the golf-related business.

This study focuses on golf course reservations, a topic that has rarely been addressed in previous research, and aims to propose an effective recommendation method while identifying and evaluating factors that influence recommendations. The main contribution of this paper lies in its ability to provide explainable recommendations, which are valuable from a practical standpoint. Unlike commodities, golf course reservations involve facilities with unique characteristics, and the decision-making process varies depending on the user's environment and preferences. Therefore, conducting new research in this area is considered to be of significant value.

This paper is structured as follows: the Introduction provides background and objectives; Dataset Overview describes the data; Analytical Methods explain the analysis techniques; Analysis Procedure outlines the recommendation system; Overall Results discuss results and promotion keyword gaps; Discussions interpret the findings; and Conclusions summarize the study and future directions.

2. Overview of Dataset

This study uses the Rakuten GORA (golf-related web service site) dataset, provided via the National Institute of Informatics by Rakuten Group, Inc [18]. This dataset includes facility data for golf courses (1,669 facilities) and review data (approximately 320,000 reviews) from the Rakuten GORA site. The facility data for golf courses contains information such as "Golf Course ID," "Golf Course Name," and "Postal Code." The review data includes elements such as "Review ID," "Golf Course ID," "Reviewer Name," "User Comment," "Play Date," and ratings for various categories. All ratings are evaluated on a five-point scale. These categories include "Cost Performance," " Staff Hospitality," "Course/Strategy," "Food Quality," and "Facilities." In this study, this data is used for analysis.

The number of users per region included in this data is shown in Table 1. Reservation numbers are higher in the Kanto region, which includes Tokyo, and the Kinki region, which includes Osaka and Kyoto, compared to other areas. In contrast, regions such as Hokkaido, Shikoku, and Kyushu—more distant from Honshu—tend to have fewer reservations.

Table 1 Number of users per region

Region	Number of reviewers
Hokkaido region	942
Tohoku region	2216
Kanto region	48711
Chubu region	9848
Kinki region	17369
Chugoku region	1934
Shikoku region	469
Kyusyu region	3450

The number of golf courses by region is shown in Table 2, which is similar to the reservation trends in Table 1. The Kanto and Kinki regions have relatively many golf courses. However, because the area proportions of these regions are lower than their population proportions, the disparity across regions is not substantial.

Table 2 Number of golf courses per region

Region	Number of reviewers
Hokkaido region	98
Tohoku region	119
Kanto region	512
Chubu region	288
Kinki region	299
Chugoku region	99
Shikoku region	43
Kyusyu region	189

3. Analytical Methods

This section describes the analytical methods used in this study. First, it introduces Word2Vec, which captures semantic relationships between words through distributed vector representations learned from review texts. Next, it explains TF-IDF, a technique to quantify the importance of words within a document set. By combining these two methods, the study integrates

semantic richness with contextual relevance for improved feature extraction. Finally, recent advances such as Transformer-based models are discussed, along with the rationale for choosing the methods employed here to balance accuracy and interpretability.

3.1. Word2Vec

Word2Vec is a natural language processing method that embeds words into a high-dimensional vector space. This method models semantic similarities and relationships between words as numerical vectors. Word2Vec has two main training models: the Skip-gram model and the continuous bag of words model (CBOW). The Skip-gram model is used to predict context words surrounding a target word and is particularly effective in capturing semantic information for rare words. In contrast, the CBOW model predicts the target word from surrounding context words, allowing faster learning across the entire dataset. These vectors represent both semantic and grammatical information, allowing for similarity searches and vector-based relational analysis. In this study, Word2Vec was utilized to learn distributed representations of words and capture their semantic and grammatical relationships from review data.

3.2. TF-IDF

Term frequency-inverse document frequency (TF-IDF) is a method for quantitatively evaluating the importance of each word in a set of documents and is widely used in the fields of natural language processing and information retrieval [19]. TF-IDF extracts important words by considering both their frequency in a document and their rarity across documents.

Term Frequency (TF) refers to the frequency of a word's occurrence in a specific document. The TF of a word t in a document d is defined by the following equation

$$TF(t,d) = term\ count\ of\ \ t\ in\ document\ d \tag{1}$$

Inverse Document Frequency (IDF) is a measure used to evaluate the rarity of a word within a collection of documents. When a word appears in many documents, its importance is reduced. IDF can be calculated by

$$IDF(t) = \log \frac{N}{df(t)} \tag{2}$$

where N is the number of all documents and $df(t)$ is the number of documents in which the word t occurs.

The TF-IDF value of a word t in document d is the product of TF and IDF, which can be represented as

$$TF\text{-}IDF(t) = TF(t,d) \times IDF(t) \tag{3}$$

By this calculation, words that occur frequently in a particular document and are rare in the whole set of documents are highly evaluated. TF-IDF is especially effective in information retrieval, text classification, and document feature extraction.

3.3. Combining Word2Vec and TF-IDF

While Word2Vec captures semantic and grammatical relationships between words through vector embeddings, it does not inherently consider the importance or relevance of words within specific documents. On the other hand, TF-IDF quantifies the importance of each word in a document set, but it does not account for semantic similarity.

Combining these two methods leverages both semantic richness and contextual relevance. Specifically, TF-IDF can be used to weight the word vectors obtained from Word2Vec, enabling the model to emphasize semantically meaningful words

that are also contextually important in each document. This combination improves the ability to represent user preferences and document characteristics more accurately, enhancing the performance of recommendation systems based on textual reviews.

Recently, Transformer models—foundational technology behind generative AI—have been applied to recommendation systems [20-21]. These methods are capable of constructing complex, high-dimensional representations, and thus, it is conceivable that such approaches can be applied to the data used in this study. However, due to the lack of pre-trained models specifically tailored for golf reservation data and the limited size of the available dataset, training such models is not feasible. Moreover, these methods are generally considered to be black boxes, which makes it difficult to derive interpretable insights such as the rationale behind individual recommendations.

Therefore, this study aims to improve recommendation accuracy by combining existing methods in a way that balances performance, interpretability, and practical applicability. By leveraging the semantic expressiveness of Word2Vec and the contextual weighting of TF-IDF, the proposed approach seeks to capture user preferences more precisely from review texts. This combination not only enhances the recommendation system's effectiveness but also maintains transparency in the model's reasoning, making it suitable for practical deployment in the golf course reservation domain.

4. Analysis Procedure

This section outlines the detailed procedure of the analysis conducted in this study. It describes the step-by-step workflow from data preprocessing and user characteristic extraction to the recommendation algorithm and evaluation method. Fig. 4 visually summarizes the overall recommendation process employed. Each subsection provides a clear explanation of the methodology used to transform raw review data into personalized golf course recommendations.

4.1. Preprocessing

(1) Scoring the Golf Courses.
- Golf courses are selected if they have received reviews from at least 10 users.
- The score for each evaluation criterion is calculated as the average of user ratings.
- The evaluation criteria consist of the following five categories: "Cost Performance," "Staff Hospitality," "Course/Strategy," "Food Quality," and "Facilities."

(2) To identify each user's characteristics, all reviews written by the same user are consolidated into a single dataset.

4.2. Extracting User Characteristics from Reviews

(1) Word2Vec was applied to vectorize user reviews.
- In this study, the Skip-gram model was adopted.
- Words were limited to nouns, verbs, and adjectives.

(2) Words related to the evaluation items and similar to the initial keywords were identified using cosine similarity.

(3) Some initial keywords were defined based on prior knowledge.

(4) Important words were extracted from user reviews using TF-IDF.

(5) The important words from the reviews were assigned to the relevant evaluation items, and the TF-IDF values were used as the score for each evaluation item.

The initial keyword lists for each evaluation item are shown below.

Cost Performance:

["cost," "performance," "fee," "price," "cost-effectiveness," "value," "reasonable," "low cost," "benefit," "cost performance"]

Staff Service:

[“staff,” “service,” “customer service,” “response,” “kind,” “polite,” “friendly,” “courtesy,” “hospitality,” “smile”]

Course/Strategy:

[“course,” “strategy,” “fairway,” “design,” “difficulty,” “challenge,” “hole,” “landscape,” “layout,” “green”]

Food Quality:

[“food,” “dish,” “taste,” “restaurant,” “menu,” “delicious,” “meal,” “foodstuff,” “gourmet,” “seasoning”]

Facilities:

[“facility,” “amenities,” “environment,” “clubhouse,” “changing room,” “cleanliness,” “well-equipped,” “comfortable,” “convenience,” “amenities”]

4.3. Recommendation

(1) The weighted sum of the evaluation item scores and the golf course rating is calculated.
(2) To account for the geographical relationship, the probability of the user having visited each prefecture is used as a weight.
(3) Multiply the weighted sum score by the prefecture weight to obtain the overall score.
(4) The top 10 golf courses are recommended based on the overall score.

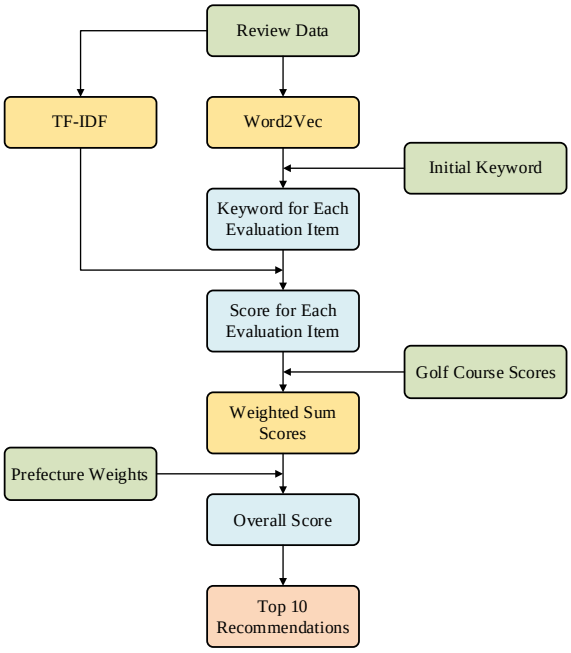

Fig. 4 Process to Recommendation

4.4. Evaluation Method

(1) Golf courses where the user has given a rating of 4 or higher and has recently visited are extracted.
(2) Identify golf courses that are geographically close to the extracted courses and have similar evaluation items, and include these courses in the “correct answer” set. A total of 10 golf courses will be selected, which are likely to satisfy the user.
(3) If a recommended golf course is included in the “correct answer” set, it is considered “correct.”
(4) Calculate the accuracy rate based on the number of “correct” recommendations.

5. Overall Results

This section presents the overall results of the recommendation system evaluation. Section 5.1 summarizes the recommendation accuracy obtained when applying the proposed method to a sample of users, highlighting the significant improvement achieved by incorporating geographic information. Next, Section 5.2 analyzes the characteristics of golf courses

most frequently recommended according to different user priorities, including “Cost Performance,” “Staff Hospitality,” “Course/Strategy,” “Food Quality,” and “Facilities.” Detailed comments and related keywords from user reviews for each category are also discussed to provide qualitative insights into the recommendations.

5.1. Recommendation Results

Table 3 shows the results of the recommendations made to 10% of all users. When the top 10 predicted scores obtained by the method used in this study were recommended to each user, the percentage of golf courses that were included in the recommendation list that the user had reserved was 23.50%.

Table 3 Recommendation results (reviews and weights of prefectures)

Correct	1643 users
Incorrect	5348 users
Percentage of correct answers	23.50%

Table 4 shows the results when user characteristics were extracted from the reviews without accounting for prefecture weights. The actual reservation rate for the top 10 was only 4.85%, which shows that the accuracy rate when using reviews alone is much lower than that of the method in this paper.

Table 4 Recommendation results (only reviews)

Correct	339 users
Incorrect	6652 users
Percentage of correct answers	4.85%

These results indicate that the accuracy of recommendations is significantly reduced when the geographic relationship between the user and the golf course is not considered. Therefore, it can be concluded that geographic information is essential for making appropriate recommendations in facility-based services such as golf course reservations.

5.2. Analysis of Correctly Recommended Results

The golf courses most frequently recommended to users prioritizing “Cost Performance,” “Staff Hospitality,” “Course/Strategy,” “Food Quality,” and “Facilities” were selected as targets for analysis. Below is a summary of the thoughts on each item.

5.2.1. Cost Performance

Below is a comment from the golf course most frequently recommended to users prioritizing “Cost Performance”:

(1) The fully separate 18 holes each bring a unique character, offering a highly strategic and engaging challenge for golfers. The fairways are flat and wide, while the greens, made of bentgrass, boast an average size of 800 square meters.
(2) The course is magnificently designed, providing a sense of freedom and exhilaration as players enjoy the game amidst the beauty of the changing seasons.
(3) This setting creates a truly relaxing and refreshing experience on the course.

The following words related to “Cost Performance” were used in the reviews of users who were recommended this course:

“reasonable,” “discount,” “premium,” “think,” “low cost,” “price,” “cost performance,” “fee,” “plan,” “get,” “overall,” “except,” “cost,” “normal,” “expense”.

The course comments did not include any of the “Cost Performance” words used by users.

5.2.2. Staff Hospitality

The following is a comment from the golf course that was most recommended to users who prioritize “Staff Hospitality”:

(1) The golf course, surrounded by ancient trees such as Sambu cedar and oak, which are over 300 years old, offers a rich natural environment. With its lush greenery, abundant water, and flat terrain, the course beautifully integrates nature.
(2) The clubhouse, with its refined French traditional style and vibrant colors, is perfectly suited for welcoming high-level executives.
(3) Located just over 50 minutes from the city center, it offers a refreshing experience akin to enjoying a forest bath.

The following words were used related to “Staff Hospitality” used by users who were recommended this course based on their reviews:

“kind,” “courteous,” “smile,” “staff,” “service,” “hospitality,” “employee,” “drink,” “reception,” “affable,” “likeable,” “attitude,” “happy,” “complimentary,” “cheerful.”

The comments for this golf course did not include any words related to “Staff Hospitality” that were used by users in their reviews.

5.2.3. Course/Strategy

The following is a comment from the golf course that was most recommended to users who prioritize “Course/Strategy”:

(1) Items such as the clubs used by J. Nicklaus during his major wins, the pin flag used in the Open Championship, and gifts from St. Andrews are currently on display. Please take this opportunity to visit and see them.

Course Introduction:

Old Course: A project team from St. Andrews was invited to supervise the design, faithfully recreating the Scottish style.

(1) "Sin Valley": A recreation of the iconic 18th hole of the Old Course, known as the "Valley of Sin."
(2) "Tommy Bunker": A replica of the "Nakajima Bunker" from the 17th hole of the Old Course.
(3) "Roman Bridge": Built using traditional methods, mirroring the Swilcan Bridge of the 18th hole of the Old Course.

New Course: Designed by J. Nicklaus and D. Muirhead, later deemed too difficult by Nicklaus himself. It is one of the most challenging courses in the country.

(1) "Victory": Tom Weisskopf once remarked, "If you finish this hole in Par for four days, you'll win the tournament."
(2) "Selection": The shaft manufacturer GRAFALLOY chose this course as part of their dream course project, selecting it from Japan.
(3) "Incident": Due to its difficulty, a maximum of 16 groups had to wait at the course's opening, which became known as the "Par 3 Incident" due to the high demand.

The following words were used related to “Course/Strategy” used by users who were recommended this course based on their reviews:

“challenge,” “tricky,” “strategy,” “course,” “good,” “think,” “go,” “round,” “enjoy,” “golf course,” “bunker,” “interesting,” “revenge,” “pond,” “fast,” “effective,” “fairway,” “change,” “slope.”

The comments for this golf course contain the words related to “Course/Strategy” used by users, which are “course” and “bunker.”

5.2.4. Food Quality

The following is a comment from the golf course that was most recommended to users who prioritize “Food Quality”:

(1) The golf course, located just 10 minutes from the Ota Kiryu IC on the Kita-Kanto Expressway, is designed to maximize the beauty of the natural mountain landscape, offering a strategically rich course. It caters to everyone, from ladies and seniors to beginners and advanced players.
(2) Visitors can enjoy the sight of 200,000 azaleas in 30 different varieties, including the 300-year-old Kirishima azalea, along with other seasonal flowers.
(3) The clubhouse, built by a major general contractor, offers a serene ambiance, inspired by an art museum.
(4) At the restaurant, guests can enjoy a meal while overlooking a Japanese garden and should try the homemade hand-pulled noodles (soba and udon).
(5) The course is committed to providing highly detailed services, such as offering amenity pouches for women, and looks forward to welcoming all guests.

The following words were used related to “Food Quality” used by users who were recommended this course based on their reviews:

“volume,” “taste,” “variety,” “meal,” “lunch,” “menu,” “dishes,” “choice,” “amount,” “cafeteria,” “items,” “rice,” “beef,” “serving,” “mass.”

The comments for this golf course include the word “meal,” which is related to “Food Quality” as used in user reviews.

5.2.5. Facilities

The following is a comment from the golf course that was most recommended to users who prioritize “Facilities”:

(1) Nasu Kasumiga-jou Golf Club has been transformed into a "true resort." The golf course, while staying true to the concept of utilizing the natural terrain created by nature, has evolved into a course loved by all golfers.
(2) For some, it offers a gentle and enjoyable resort-style course, while for others, it provides a challenging, competitive experience.
(3) The clubhouse, the largest wooden commercial facility in Japan, boasts interiors designed by Joan Behnke, an interior designer renowned for works such as MGM Grand. Inside the clubhouse, the resort hotel features four suite rooms with terrace jacuzzi baths, 30 guest rooms, two restaurants, a bar lounge, an outdoor hot spring, and a spa.
(4) The golf course has undergone layout changes and expanded fairways, and the entire course is now covered with Korean lawn grass. While maintaining its strategic challenges, the course also incorporates the gentleness of a resort. Furthermore, electromagnetic-guided carts have been introduced to promote environmental conservation. The use of wooden tees is also encouraged to promote harmony with nature, and cooperation is appreciated.

The following words were used related to “Facilities” used by users who were recommended this course based on their reviews:

“pleasant,” “luxurious,” “smooth,” “good,” “fun,” “facilities,” “hot spring,” “clubhouse,” “lockers,” “restroom,” “old,” “locker room,” “bathroom,” “cool,” “abundant,” “neat,” "dressing room,” “clean.”

The comments for this golf course include the words “hot spring,” “clubhouse,” and “facilities,” which are related to “Facilities” as used in user reviews.

6. Discussions

When geographic factors are not considered in the model, it was found that the accuracy of recommendations significantly decreases. This suggests that golf course data is greatly influenced by geographic factors. Therefore, in fields like e-commerce product reviews, where geographic factors are less influential, the proposed method of extracting user characteristics from reviews may be more effective.

In the analysis of correct recommendation results, it was found that in the categories of "Cost Performance" and "Staff Hospitality," the golf course comments often lacked the specific keywords prioritized by users. Golf courses that excel in "Cost Performance" should incorporate related keywords in their promotional content. This approach could help attract new users who prioritize "Cost Performance" and prevent user churn. Likewise, those excelling in "Staff Hospitality" should include relevant terms to attract new users and reduce churn.

In contrast, for "Course/Strategy," "Food Quality," and "Facilities," the recommended courses' comments did include relevant words. However, only up to three words from each category were present. Emphasizing these words more in promotions may help bridge the gap between user expectations and course descriptions. This approach could lead to the acquisition of more new users and a decrease in user churn.

To evaluate the effectiveness of the proposed method, it was compared with several other approaches. The comparison methods included the following: (1) recommendation scores based only on Word2Vec, (2) scores based only on regional information, and (3) a combination of Word2Vec and collaborative filtering. The results are summarized in Table 5.

As shown in Table 5, the proposed method achieved the highest accuracy, surpassing even the combination with collaborative filtering, which is generally considered to yield high performance. The method using only regional data also showed high accuracy, likely because golf reservations are tied to users' residential areas.

Table 5 Evaluation of the method of this study

Method	No. of Correct	Correct Percentage
(1) Used only on Word2Vec	320	4.58%
(2) Used only on regional information	1530	21.89%
(3) Combination of Word2Vec and collaborative filtering	1330	19.02%
(4) Combination of Word2Vec and TF-IDF (method of this study)	1542	22.06%

These findings indicate that capturing user preferences from reviews outperforms collaborative filtering that focuses solely on course features. In particular, the accuracy rate is much higher than when only reviews are used. Although reviews are known to influence decision-making, this suggests that using reviews alone is not enough to make good recommendations.

The method of this study achieved an accuracy rate of approximately 22%. To the best of current knowledge, no previous studies have focused primarily on golf course recommendations. However, However, the results can be compared with those reported by Traub et al., who addressed hotel recommendation—a comparable task involving facility reservations [22]. In that study, multiple recommendation techniques were combined with location-based filtering, similar to the approach in this study. Nevertheless, the reported accuracy was approximately 4%, which is significantly lower than ours. Although a direct comparison is difficult due to the much larger number of hotels compared to golf courses, this contrast still suggests that the accuracy achieved in this study is reasonably high.

Considering that golf course selection is influenced by a wide range of factors, such as course difficulty, available facilities, service quality, and geographical conditions, the accuracy achieved by the method proposed in this article is by no means low and can be regarded as sufficiently high.

7. Conclusions

This study proposed a recommendation method tailored to golf course users by leveraging textual reviews and incorporating geographic information. The methodology aimed to identify user preferences using TF-IDF and Word2Vec, and to match those preferences with golf course features for more accurate and interpretable recommendations. The process further accounted for users' visiting history to adjust for geographical tendencies, enabling the system to generate personalized and practical recommendations.

The major findings of this study are as follows:

(1) Recommendation Accuracy: The proposed method achieved a correctness rate of 23.50%, significantly outperforming the review-only approach (4.85%) and even other combinations such as collaborative filtering with Word2Vec. This highlights the effectiveness of incorporating both textual preferences and geographic data.
(2) Role of Geographic Information: Geographic proximity had a substantial impact on recommendation accuracy. Without it, the accuracy dropped sharply, confirming that golf course selection is highly location-dependent.
(3) Keyword Gaps in Promotions: In categories such as "Cost Performance" and "Staff Hospitality," there was a notable gap between the vocabulary used by users and the promotional language in golf course descriptions. In contrast, courses that matched user preferences in "Course/Strategy," "Food Quality," and "Facilities" showed partial alignment in keyword usage. These findings suggest that promotional effectiveness could be improved by incorporating review-based keywords.
(4) Interpretability vs. Complexity: While modern Transformer-based models can offer higher representational capacity, they were not employed in this study due to data limitations and interpretability concerns. Instead, the combination of TF-IDF and Word2Vec provided a balanced approach that supports both performance and transparency.

Given that golf course reservations are influenced by highly individual factors such as personal taste, location, and service quality, the achieved accuracy is considered meaningful even if the numerical value appears modest when compared with traditional product recommendations.

Future directions include enhancing geographic modeling by incorporating transportation accessibility (e.g., proximity to interchanges) and testing whether integrating user-centric keywords into promotional materials improves user engagement and reservation rates. These extensions will further bridge the gap between user expectations and facility-side communication strategies.

Acknowledgments

In this paper, the "Rakuten Dataset" (https://rit.rakuten.com/data_release/) provided by Rakuten Group, Inc. via the IDR Dataset Service of the National Institute of Informatics was used. This work was supported by JSPS KAKENHI Grant Numbers 24H00370

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Ministry of Internal Affairs and Communications, "2016 Information and Communications White Paper," https://www.ituaj.jp/wp-content/uploads/2016/10/nb28-4_web-07-Report-Overview2016WhitePaper.pdf, 2016.
[2] Ministry of Economy, Trade and Industry, "Long-Term Data," https://www.meti.go.jp/statistics/tyo/tokusabido/result/result_1.html, 2024.
[3] M.A.H. Bekken, P.D. Mitchell, and D.J. Soldat, "An Eco-Efficiency Model for Golf," Ecological Indicators, vol. 166,

article no. 112357, 2024.
[4] Y. Izumi, H. Shigeoka, and M. Yagasaki, “Golfing CEOs,” Labour Economics, vol. 91, article no. 102639, 2024.
[5] N. Edwards, C. Dickin, & H. Wang, “Low Back Pain and Golf: A Review of Biomechanical Risk Factors,” Sports Medicine and Health Science, vol. 2, no. 1, pp. 10-18, 2020.
[6] M. Khazaeli, ' L. Javadpour, “Golf Club Selection with AI-Based Game Planning,” Entropy, vol. 26, no. 9, article no. 800, 2024.
[7] H. Morise, K. Atarashi, S. Oyama, and M. Kurihara, “Neural Collaborative Filtering with Multicriteria Evaluation Data,” Applied Soft Computing, vol. 119, article no. 108548, 2022.
[8] E. Asani, H. Vahdat-Nejad, and J. Sadri, “Restaurant Recommender System Based on Sentiment Analysis,” Machine Learning with Applications, vol. 6, article no. 100114, 2021.
[9] M. Elahi, D. K. Kholgh, M. S. Kiarostami, M. Oussalah, and S. Saghari, “Hybrid Recommendation by Incorporating the Sentiment of Product Reviews,” Information Sciences, vol. 625, pp. 738-756, 2023.
[10] Z. Abbasi-Moud, H. Vahdat-Nejad, and J. Sadri, “Tourism Recommendation System Based on Semantic Clustering and Sentiment Analysis,” Expert Systems with Applications, vol. 167, article no. 114324, 2021.
[11] I. Syafganti, and M. Walrave, “Assessing the Effects of Valence and Reviewers’ Expertise on Consumers’ Intention to Book and Recommend a Hotel,” International Journal of Hospitality & Tourism Administration, vol. 23, no. 5, pp. 904-923, 2022.
[12] X. Wang, J. Zheng, L. Tang, and Y. Luo, “Recommend or Not? The Influence of Emotions on Passengers′ Intention of Airline Recommendation during COVID-19,” Tourism Management, vol. 95, article no. 104675, 2023.
[13] J. Zhou, Z. Ye, S. Zhang, Z. Geng, N. Han, and T. Yang, “Investigating Response Behavior through TF-IDF and Word2Vec Text Analysis: A Case Study of PISA 2012 Problem-Solving Process Data,” Heliyon, vol. 10, no. 16, article no. e35945, 2024.
[14] Z. Zhang, F. Han, H. Zhang, T. Aoki, and K. Ogasawara, “Examining the Effect of the Ratio of Biomedical Domain to General Domain Data in Corpus in Biomedical Literature Mining,” Applied Sciences, vol. 12, no. 1, article no. 154, 2022.
[15] A. Al Tawil, L. Almazaydeh, D. Qawasmeh, B. Qawasmeh, M. Alshinwan, and K. Elleithy, “Comparative Analysis of Machine Learning Algorithms for Email Phishing Detection Using TF-IDF, Word2Vec, and BERT,” Computers, Materials and Continua, vol. 81, no. 2, pp. 3395-3412, 2024.
[16] A. Liyih, S. Anagaw, M. Yibeyin, and Y. Tehone, “Sentiment Analysis of the Hamas-Israel War on YouTube Comments Using Deep Learning,” Scientific Reports, vol. 14, no. 1, article no. 13647, 2024.
[17] M. A. Abid, M. F. Mushtaq, U. Akram, M. A. Abbasi, and F. Rustam, “Comparative Analysis of TF-IDF and Loglikelihood Method for Keywords Extraction of Twitter Data,” Mehran University Research Journal of Engineering and Technology, vol. 42, no. 1, pp. 88-94, 2023.
[18] Rakuten Group, Inc., “Rakuten GORA Data,” https://dsc.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=1754&item_no=1&page_id=13&block_id=21, 2010.
[19] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.
[20] A. Subbiah, and V. Aggarw, “Transformers in Music Recommendation”, https://research.google/blog/transformers-in-music-recommendation/, 2024.
[21] G. de Souza Pereira Moreira, S. Rabhi, J. M. Lee, R. Ak, and E. Oldridge, “Transformers4Rec: Bridging the Gap between NLP and Sequential / Session-Based Recommendation,” Proceedings of the 15th ACM Conference on Recommender Systems (RecSys’21), pp. 143-153, 2021.
[22] M. Traub, D. Kowald, E. Lacic, P. Schoen, G. Supp, and E. Lex, “Smart Booking without Looking: Providing Hotel Recommendations in the TripRebel Portal,” Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business, article no. 50, pp. 1-4, 2015.