International Journal of Mathematical, Engineering and Management Sciences
Vol. 11, No. 4, 1535-1559, 2026
https://doi.org/10.33889/IJMEMS.2026.11.4.063

Beyond Prediction: Using Explainable AI to Uncover the Drivers of Customer Ratings with SHAP and Review Topics

Sule Ozturk Birim
Department of Business Administration,
Salihli Faculty of Economics and Administrative Sciences,
Manisa Celal Bayar University, Manisa, Turkiye.
E-mail: sule.ozturk@cbu.edu.tr

Ipek Kazancoglu
Department of Business Administration,
Faculty of Economics and Administrative Sciences, Ege University, Izmir, Turkiye.
Corresponding author: ipek.savasci@ege.edu.tr

Yigit Kazancoglu
Logistics Management Department,
Yasar University, Universite Cad No.37-39, Bornova, Izmir, Turkiye.
E-mail: yigit.kazancoglu@yasar.edu.tr

(Received on January 18, 2026; Revised on March 24, 2026 & May 18, 2026 & June 9, 2026; Accepted on June 18, 2026)

Abstract
Understanding the drivers of customer satisfaction is significant for luxury restaurants to maintain and increase consumers. While online reviews provide a rich source of customer opinions, extracting clear and actionable insights from this unstructured text carries a major difficulty for decision makers. This paper proposes a framework to not only predict customer star ratings from reviews, but also to explain the key drivers behind those predictions. BERTopic is applied on a dataset of reviews from Michelin-starred restaurants in Turkey. Four different feature engineering strategies are compared in this study, changing based on the integration of topic modeling and sentiment analysis. Five machine learning models were leveraged to predict the star rating, and the SHAP framework was used to interpret best-performing models and select best drivers of prediction. The results show that feature sets which include sentiment outperform those based on only topic presence. Weighted sentiments feature achieved the best predictive performance, with the XGBoost yielding the lowest error. The SHAP analysis revealed that the model's predictions are mainly driven by primary performance factors like Dining Experience and Staff Service. Staff Service and Reservation Process were found as features for which a failure strongly lowers the rating, while a success provides almost no benefit on ratings. This research provides a contribution with successfully identifying and quantifying the key drivers of customer satisfaction. By combining topic modeling with explainable AI, this work moves beyond prediction to provide actionable insights.

Keywords- Explainable AI, SHAP, Topic modeling, Rating prediction, Customer satisfaction.

1. Introduction

As information is shared rapidly between customers digitally, understanding and improving customer satisfaction has become essential for businesses. This evolving environment has essentially changed how customers make decisions about a service or a product. Online review platforms, like TripAdvisor have become a critical source of information where customers share their experiences in their own words which are User Generated Content (UGC). UGC has a strong influence on the reputation of a restaurant and the booking intentions of future customers (Lu & Stepchenkova, 2015). For businesses, these reviews represent a large and valuable source of unstructured feedback that contains detailed information about what is appreciated or disapproved by the customers.

This situation is especially relevant in the luxury dining sector, such as restaurants that have been awarded Michelin stars. These establishments operate under very high customer expectations and their success depends on delivering superior experience (Lee & Hwang, 2011). A single negative review can have a strong negative impact on the restaurant's reputation. Therefore, for managers of these luxury restaurants it's not sufficient to simply know their average star rating from UGC. They need to understand the specific drivers behind those ratings such that what specific aspects of the food, service or atmosphere led to a five-star review or what specific failures led to a low star review. Answering these questions is the primary motivation for this study. The goal is to develop an analytical framework that can deconstruct customer reviews from Michelin-starred restaurants to provide clear data-driven understanding of the core drivers of customer satisfaction.

The main challenge, however, is that this valuable information is hidden within large volumes of unstructured text. To analyze this data researchers often use a combination of two powerful computational techniques, topic modeling and sentiment analysis. Topic modeling is used to automatically discover what customers are talking about, by identifying the main themes or topics in the text (Tirunillai & Tellis, 2014). Sentiment Analysis, on the other hand, helps to understand how customers feel about those topics by measuring the positive or negative emotion in their language (Namkung & Jang, 2008).

While the value of online reviews is obvious, the best method for using them to get predictive insights is still an area open for research. Previous studies have mostly used topic modeling and sentiment analysis separately (Namkung & Jang, 2008; Tirunillai & Tellis, 2014), and this has left some important gaps in the literature, particularly in the context of luxury dining where customer expectations are very high and predictive studies remain scarce. First, it is still not clear which type of information is more useful for predicting star ratings: what customers talk about (topic proportions) or how they feel about those topics (topic sentiment). Second, very few studies have tested whether combining both dimensions into a single feature can produce a stronger prediction than using either one alone (Birim & Kazancoglu, 2026). Third, most existing models do not explain their predictions in a way that is useful for managers, limiting their practical value for decision making. To the best of our knowledge, this study's novelty lies in evaluating topic proportions, topic sentiment, and their combination as predictive features for customer star ratings in Michelin-starred restaurants, while also applying explainable AI to interpret the results.

To address the gaps and guide our research, the following research questions are formulated:

(i) Can we extract powerful predictive features from the review text using topic modeling?
(ii) Which is more predictive: what customers talk about (topic proportions) or how they talk about it (topic sentiment)?
(iii) How does a feature that combines both proportion and sentiment perform in predicting reviewer satisfaction?
(iv) Which features have the most influence on satisfaction prediction?

To answer these four research questions, this paper represents a comprehensive study that involves several steps. First, topic modeling is used to extract hidden topics from a corpus of reviews of Michelin-starred restaurants in Turkey. Then the study engineers four different sets of features: document-level topic probabilities, sentence-level topic proportions, average topic sentiments, and weighted topic sentiments. Five different machine learning models were compared on these four feature sets to find the most accurate combination. Finally, the Shapley Additive Explanations (SHAP) framework is applied to the best performing models to interpret their predictions and uncover the key drivers of high and low ratings. This study aims to provide a clear methodological framework for researchers and actionable insights for managers in the fine-dining industry. Specifically, this study makes the following contributions:

- It proposes and compares four different feature sets based on topic modeling, which allows a direct comparison of different ways to represent review content for prediction purposes.
- It directly tests whether topical focus, emotional tone, or a combination of both is the most useful predictor of customer ratings, which is a question that has not been clearly answered in the literature.
- It applies the SHAP framework to explain model predictions and identify the specific aspects of the dining experience as the most influential variable on customer satisfaction in Michelin-starred restaurants.
- It provides a clear and replicable analytical framework that is tested in the fine-dining sector in Turkey, which is an under-studied context in the hospitality literature.

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature on online reviews, customer satisfaction, topic, modeling and the use of explainable AI in customer rating prediction. Section 3 describes the full methodology from the data collection to feature engineering and predicting modeling. Section 4 presents the results of the model comparisons and detailed insights from the SHAP analysis. Finally, Section 5 discusses the theoretical and practical implications of the findings; Section 6 concludes the paper with limitations and directions for future research.

2. Literature Review

This section reviews the existing literature on three key areas that inform this study. Customer online reviews, customer satisfaction for Michelin starred restaurants and topic modeling and Shapley Additive Explanations (SHAP)in the context of hospitality and service quality research.

2.1 Customer Online Reviews

The increasing use of the internet and social media platforms affects consumers to consider online reviews before making purchasing decisions (Park et al., 2020). In the hospitality sector, online reviews influence consumers as a source of information when choosing restaurants due to the intangible nature of service (Rita et al., 2023). In this way, consumers share their opinions, comments, and experiences on online platforms that increase the trust of other consumers, reduce search of information time, and decrease the risk of purchase. UGC generates a large amount of information (Gallinucci et al., 2015; Tirunillai & Tellis, 2014) and creates an electronic word-of-mouth (e-WOM) effect by influencing other consumers (Lu & Stepchenkova, 2015; Park et al., 2020; Tsao & Hsieh, 2015). There are online review platforms for retailer, restaurant, and hotel reviews such as TripAdvisor, Zomato, Yelp, or Google Reviews. On these platforms, online reviews generally consist of quantitative ratings, usually scored on a scale of 1 to 5 stars, or qualitative reviews based on written comments (Rita et al., 2023). Consumers share their restaurant experiences and recommendations on these platforms, expressing their feelings (Xu, 2021). This feeling expression in the UGC can help restaurant managers to monitor service performance, understand customer perceptions or intentions, and develop new products or services (Bang et al., 2022). For this reason, understanding online reviews is necessary for restaurant owners to create a competitive advantage, gain insight into customer attitudes and behaviors, and improve service quality (Luo & Xu, 2021).

2.2 Customer Satisfaction for Michelin Starred Restaurants

The Michelin Guide is one of the most prestigious certifications and authoritative indicators in the global gastronomy industry. Michelin-starred restaurants are generally positioned in the luxury segment as fine dining establishments (Pacheco, 2018), and are defined as haute cuisine (Svejenova et al., 2007). In this respect, the Michelin Guide is considered the international benchmark for cuisine quality (Chiang & Guo, 2021; Harrington et al., 2013; Michelin Guide, 2022). Michelin stars are not only an award in the restaurant industry, but also serve to increase the number of customers and promote customer loyalty (Harrington et al., 2013), create an attraction for tourism destinations (Castillo-Manzano et al., 2021; Daries et al., 2021),

and are used as a marketing tool to enhance the reputation and visibility of restaurants (Bang et al., 2022; Farrugia, 2024). Furthermore, restaurants continuously innovate and strive for creativity, improving customer experiences in order to obtain and maintain their Michelin stars (Bang et al., 2022; Martins, 2022). Michelin-starred restaurants are evaluated according to five criteria: the quality of ingredients used, the chefs' mastery of flavor and cooking techniques, the chef's personality, the value of the food, and the level of consistency in both the menu and throughout the year (Ho et al., 2021; Michelin Guide, 2022). The Michelin Guide rating system consists of a maximum of 3 stars. One star represents a restaurant with high-quality food preparation, evaluating only the quality of food and wine. Two stars reflect excellent culinary skills, originality in the dishes, and the chef's personality, making it worth a trip. Three stars represent a restaurant with exceptional cuisine, where customers frequently enjoy excellent meals, an elegant dining room, and excellent service, making perfect the total experience (Farrugia, 2024; Martins, 2022). Several factors influence customer satisfaction such as food quality, staff behavior, menu variety, restaurant location, atmosphere, ambiance, price level, perceived value, cleanliness, reputation, and service quality (Choi et al., 2021; Luo & Xu, 2021; Kim & Hwang, 2022). Among these factors, food quality (Namkung & Jang, 2008; Pantelidis, 2010; Sulek & Hensley, 2004), employee performance, atmosphere, and perceived food taste level are seen to play a significant role in customer satisfaction (Bertan, 2016).

Studies on customer satisfaction in Michelin-starred restaurants have revealed that various factors (high-quality service, delicious food, magnificent physical environment, unique atmosphere, and high perceived value) influence customer satisfaction (Chang et al., 2025; Michelin Guide, 2022). Michelin-starred restaurants are positioned as part of the luxury segment (Pacheco, 2018). In this context, consumers have higher expectations at luxury restaurants, expecting not only good food, but also higher quality service, a unique and exclusive atmosphere and experience (Berry et al., 2006; Barrera-Barrera, 2023; Rita et al., 2023; Yang & Mattila, 2016), paying attention to the appearance of the staff and the service. These factors have an influence on customer satisfaction and loyalty (Amelia & Garg, 2016; Chaturvedi et al., 2024). At the same time, consumers are also willing to pay high prices for these restaurants (Lee & Hwang, 2011). Furthermore, positive experiences, high-quality food, staff behavior, and reasonable prices are also emphasized as factors that increase customer satisfaction (Reis, 2024). Factors regarding the restaurant atmosphere, such as decoration, location, appearance, music, furniture, scent, lighting, temperature, and building structure, influence the customer's mood and positively increase their satisfaction (Barrera-Barrera, 2023; Liu et al., 2015; Namkung & Jang, 2008). Therefore, while the literature provides a comprehensive list of what contributes to a luxury dining experience, this study seeks to understand the relative importance of these different drivers from the customer's own perspective, as expressed in their reviews.

Previous studies on Michelin-starred restaurants have generally adopted traditional research methods such as surveys and interviews to understand consumer perceptions (Chang et al., 2025; Chen & Lin, 2025; Liu et al., 2015; Pacheco, 2018; Tsaur & Lo, 2020). However, in recent years, online review platforms, which have become significant in consumers' decision-making processes, have caused academic studies to adopt text-based analytical methods. Online reviews are a very good source for examining customer satisfaction of restaurants because they do not contain sampling bias compared to surveys (Schuckert et al., 2018). Online reviews provide valuable insights into customer opinions, experience, and emotional responses beyond numerical ratings (Temiz et al., 2025). Given this, consumers are influenced by the information and emotions expressed in online reviews. Consumers share positive emotions such as excitement and pleasure or negative emotions such as anger, disappointment, and resentment in relation to their restaurant experience in online reviews (Kim et al., 2025). Analyzing customer emotions provides critical insights into understanding satisfaction levels and improving the restaurant experience. However, there are limited studies investigating customer satisfaction at Michelin restaurants based on online customer reviews

(Barrera-Barrera, 2023; Kim et al., 2024; Nguyen & Ho, 2023; Rita et al., 2023; Sangkaew et al., 2025; Temiz et al., 2025). To address these gaps, this study introduces a multi-stage analytical framework. First, topic modeling is applied to identify what customers discuss, and sentiment analysis is used to capture how they feel about those topics. These two dimensions are then combined into a unified feature set for predicting customer ratings. Finally, explainable AI is applied to identify the most influential drivers of customer satisfaction.

2.3 Shapley Additive Explanations (SHAP)

SHAP is a model-agnostic, post-hoc explanation method from the field of explainable Artificial Intelligence (XAI) that is based on the Shapley values from game theory framework. SHAP assigns each predictor an importance value for a particular task by testing how the model's prediction changes when a single feature is added to every possible combination of other features. Then the final importance of a feature is calculated as taking the average of its importance value across all the combinations (Lundberg & Lee, 2017). SHAP is a post-hoc interpretation method that can be applied to explain the output of any machine learning model. In this study, SHAP helps to understand the contribution of each feature to the prediction of star rating. A positive SHAP value means, the feature's value for that review increases the rating prediction, while a negative SHAP value means it decreases value of predicted rating.

The use of SHAP for interpreting complex machine learning models in prediction tasks related to customer reviews analysis is popularly adapted in the literature. For example, in the hospitality field, Wang et al. (2024) used SHAP values to measure the impact of different service attributes, and to better understand customer priorities. Similarly, Aparicio-Aparicio et al. (2024) applied SHAP to identify how specific product attributes in outfit industry, positively or negatively influenced online ratings. Other researchers used SHAP to understand complex patterns, such as how the effect of review length on helpfulness score is affected by other features in the text (Meng et al., 2021).

While previous studies demonstrate SHAP as a powerful tool in the analysis of UGC, the study is distinct in how the features used for prediction are created and leveraged. The approach is data driven using only customer reviews. Unlike studies by Meng et al. (2021) and Wang et al. (2024), which rely on pre-defined aspect level ratings provided by the online platform or features derived from existing theories, this study's feature sets are generated directly from the topics that are calculated using the customer review content. This study uses a neural topic modeling methodology. While previous studies have successfully used traditional models like LDA to create topics (Aparicio-Aparicio et al., 2024; Yang et al., 2024), this research uses BERTopic, a modern transformer-based model, that considers contextual meaning of the text. Additionally, four different feature sets from these topics are compared to identify the most powerful prediction model. Finally, the research differs in its use of sentiment in rating prediction. Zhang et al. (2025) used BERTopic with SHAP for the task of binary sentiment classification as positive or negative. In contrast, this research incorporates sentiment as the predictive feature to understand its contribution on rating prediction, not as the target variable. By combining advanced topic modeling approach with predictive analysis and an explainable AI framework, this study provides insights about how the features customers discuss influence their satisfaction.

3. Methodology

This section describes the methodology of the study including dataset description, topic modeling, feature engineering, predictive modeling and how the models are evaluated. A procedural analysis is conducted to uncover the drivers of customer ratings and several models were compared to find the best performing one. Details of the methodological framework is presented in **Figure 1**.

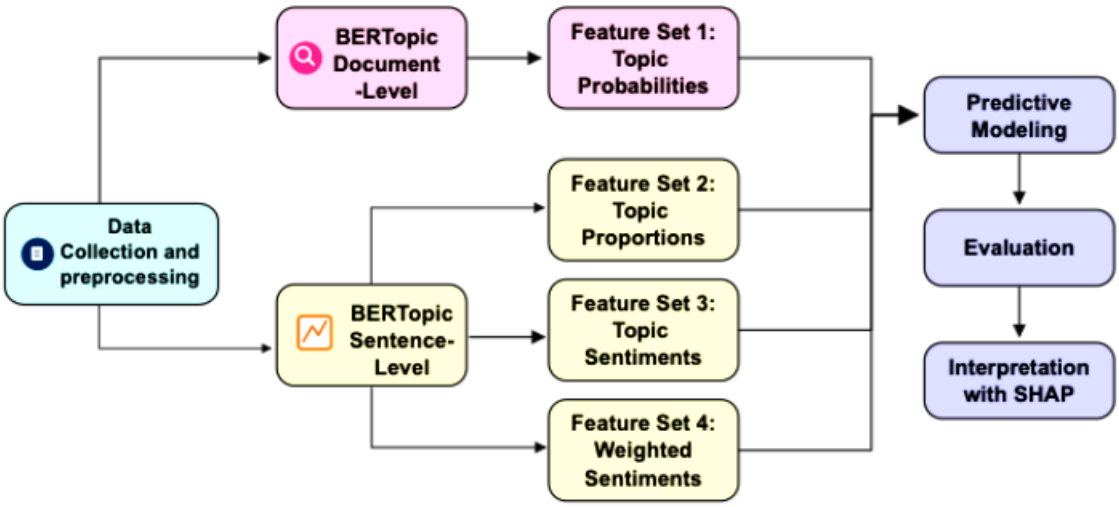


Figure 1. Methodological framework.

3.1 Data Collection and Pre-processing

The dataset for this study was collected from the popular online review platform, TripAdvisor. Initially, all reviews from the eleven restaurants that held one Michelin star at the time of data collection were collected. From the initial list of restaurants, two restaurants that are located inside the hotels were excluded from the analysis. This decision was made to create a more consistent dataset, focusing on only restaurant experiences and avoiding factors related to hotel amenities. The restaurants from which the reviews are taken are located in Izmir, Istanbul and Bodrum provinces of Turkey.

Each entry in the dataset contains two key pieces of information. First it has the full text of the customer's review and secondly it has the corresponding star rating given by that customer. The star rating is the target variable that the machine learning models will try to predict. All Turkish reviews were translated to English using Google Translate API. Translation is necessary because advanced topic modeling tools like BERTopic are primarily trained and optimized for English. Additionally, translating reviews provides a language for the analysis that is consistent with the entire paper.

To prepare the raw review texts for analysis there is a need to preprocess the data. This preprocessing is necessary to clean the text and make it suitable for topic modeling analysis. In preprocessing several steps were applied. First, all text was converted to lowercase to ensure that words like "Staff" and "staff" are treated as the same. Next, all punctuation marks, numerical digits, and other special characters were removed. The main goal of this text cleaning process is to reduce the noise in the data and create a standardized set of words to be used in topic modeling.

3.2 Topic Modeling with BERTopic

To discover the main themes discussed in the reviews, BERTopic was selected as the topic modeling technique. The choice of BERTopic over the more traditional Latent Dirichlet Allocation (LDA) was made for several reasons. First, LDA relies on word co-occurrence patterns using a Bag-of-Words approach, which ignores the order and context of words in a sentence (Blei et al., 2003). This is a significant limitation when analyzing UGC such as restaurant reviews, where meaning is often dependent on context. For

example, the word "cold" can refer to food temperature or to the attitude of the staff, and LDA cannot distinguish between these two uses. Second, LDA requires sufficient word co-occurrence across documents to estimate reliable topic distributions. In short texts like online reviews, there are simply not enough words per document to produce stable and meaningful co-occurrence patterns, which leads to incoherent topics (Vayansky & Kumar, 2020; Zhao et al., 2011). Third, BERTopic uses transformer-based word embeddings that capture the semantic meaning of words in context, which leads to more coherent and separable topics (Grootendorst, 2022). Studies in the literature indicate that using word embeddings in topic modeling makes the model to create more separable and coherent topics with relevant reviews inside the cluster (de Lima et al., 2023; Sanchez-Franco & Rey-Moreno, 2022). For these reasons, BERTopic is considered a more appropriate choice for analyzing the short, context-dependent, and informal language found in restaurant reviews.

3.3 Two-stage Approach for Topic Modeling

BERTopic is used in two different ways to analyze the review data. This two-stage approach allowed to capture both broad themes and more specific clusters that are discussed in the reviews. Two approaches are as following:

***Document-level topic modeling*:** In this approach, each review is treated as a single, whole document. This approach is used to identify broad themes in the review corpus and the main inference is a restaurant review is generally concentrated on a few broad themes (Yang et al., 2024). They are previously used to reduce a large collection of text into a distribution of topics. These models mostly assume that each document is a multinomial distribution over topics, each document has a probability value of belonging to a topic in LDA-like topic models (Jiang et al., 2019). In this study, the purpose of this approach is to identify the general themes that a review might focus on giving a broad perspective of the topics in the corpus and calculate probabilities of the topics for each document.

***Sentence-level topic modeling*:** This approach focuses on the sentence as the primary unit of processing and assume each sentence concentrates mostly on one theme (Cha & Lee, 2024; Huang et al., 2022). This approach is found to be discriminate between identical words used in different contexts (Cha & Lee, 2024). Additionally, it is meant to identify the internal structure of a text such as distinguishing between a summary sentence and a sentence including details (Huang et al., 2022). Additionally, in BERTopic documentation to find more than one topic associated with a review, it is recommended to split the reviews into sentences and then combine them to obtain a topic distribution based on the frequency of topics observed in a review (Grootendorst, 2024). To capture more granular topics within the review and to find different topics that are relevant in a review, the reviews were split into sentences and the topic model analysis was then applied as a practice observed in the literature (Amirifar et al., 2023; Xu et al., 2025).

3.4 Sentiment Analysis

After the topic modeling the next step was to calculate sentiment scores from the review texts. For the analysis sentiment is calculated at the sentence level because by calculating sentiment for each sentence, the average sentiment for each specific topic within a single review can be calculated. For example, in one review, sentiment for sentences about “Service Quality” can be calculated and compared to the sentiment for sentences related to “Meat Dishes”.

This method is different from a document-level approach, which calculates sentiment of a whole review. A single sentiment score for a review is not practical to assign to the different topics that are discussed in a long review. A customer might write positively about the food but negatively about the staff attitude in one review, and assigning a single sentiment score cannot capture this detail. Additionally, the overall sentiment

of a review often has a very direct and high correlation with the user star rating (Pang & Lee, 2008). Using it as a predictive measure could be less helpful for this study's goal, which is understanding the influence of specific dining features on the prediction of customer scores.

To perform the sentiment analysis, a sentiment model from the Hugging Face Transformers library, "nlptown/bert-base-multilingual-uncased-sentiment model" is used. This model has been pre-trained on a large corpus and finetuned for the task of sentiment analysis of product reviews (NLP Town, 2024). The high accuracy of this model for predicting sentiment in customer reviews and on social media data has been demonstrated in several recent studies (Miah et al., 2024; Sahoo et al., 2023). The model provides a sentiment score for each sentence on a scale from 1 (very negative) to 5 (very positive). These sentence-level scores were used to build the sentiment-based feature sets which are described in the next section.

3.5 Feature Engineering

After identifying the topics from the review texts, the next step is to transform this information into numerical features that a machine learning model can use for prediction. To test which aspects of a review are most predictive of its star rating four different sets of features were designed. The first set was derived from the document-level topic model, while the other three, more detailed sets, were derived from the sentence-level topic model.

***Feature Set* 1 - *Document-based probabilities*:** This set was designed to understand the main theme of the review. It is created using the output of document-level topic model. To obtain multinomial distribution over topics for one whole review, topic probabilities for each review from the BERTopic model were extracted. HDBSCAN clustering algorithm is a key component in BERTopic which not only groups similar documents into clusters but also generates a probability distribution for each document across the identified topics (Grootendorst, 2022). These probability scores indicate the strength of a document's association with a topic (McInnes et al., 2017). For every review in the dataset, a feature vector where each element is the probability score for one of the topics is created. This feature set aimed to capture the main theme along with the additional themes covering an entire review. For example, a review might have a high probability for the "Dining Experience" and low probabilities for others.

***Feature Set* 2 (*Sentence-based*)- *Topic proportions*:** This set is designed to measure the focus of the reviewer. Desire is to know what specific topics a customer dedicated the attention to within the review. To calculate this, the outputs of sentence-level topic model is used. For each review the topic of every single sentence is identified. Then for each topic the number of sentences that belonged to it are counted. Then this count is divided by the total number of sentences in that review. This approach is similar to the Attribute Distribution Model applied previously (Birim & Kazancoglu, 2026). The resulting value called "Topic Proportion" is a percentage that represents how much of the review's text is focused on the specific feature.

***Feature Set* 3 (*Sentence-based*)- *Average sentiments*:** this set is designed to capture the emotional tone associated with each specific topic. While topic proportions tell what a customer focused on, this feature set tells how they feel about it. Using sentence-level topics and scores of the previously described sentiment model, this feature is calculated. For each review, all sentences belonging to a topic are obtained, and then the average sentiment score of just those sentences are calculated to gather average sentiment for a specific topic in that review. This approach is similar to the methodology applied for aspect sentiment analysis in Shahhosseini & Nasr (2024) and Sentiment Enhanced model applied in Birim & Kazancoglu (2026). This feature set allowed to capture varying level of sentiment in a review. For example, a review could have a very positive average sentiment for the "Meat Dishes" while a very negative average sentiment for the "Wine Pairing" topic.

***Feature Set* 4 (*Sentence-based*)- *Weighted sentiments*:** This set is designed to capture the reviewer's attention to a topic by combining the focus and emotional tone regarding the topic. This feature is calculated as explained in the Weighted Sentiment Model of Birim and Kazancoglu (2026) by multiplying the topic proportions from Feature Set 2 by the average sentiment from feature set 3 for each topic in a review. The idea behind this feature set is that a single positive or negative sentiment is more meaningful if the topic was the prominent focus in the sentence. For example, a slightly negative sentiment on a topic that makes up eighty percent of a long review can be more powerful than a very negative sentiment on a topic that was only mentioned in one sentence. An emotional tone of a feature will be very strong if it is estimated sentiment has a high score and its topic proportion is also high in the review. Therefore, this feature is designed to separate high influential opinions from briefly discussed comments. The distinction of the weighted sentiments from Birim & Kazancoglu (2026), comes from the application of this approach in star rating prediction via machine learning methods and further use in an explainable AI framework. The aim is to observe if this approach could provide more powerful predictions than the other three feature set approaches.

3.6 Predictive Modeling

After the generation of feature sets, it is necessary to develop the predictive models, with the aim of predicting star ratings for each review. To apply a comprehensive and robust analysis five machine learning models were selected. This multi-model approach allows the comparison of the performances of the models and to determine which of the proposed feature sets provide consistently the highest prediction accuracy. Machine learning models were chosen to represent a diverse range of learning methods.

The first group of models included two tree-based ensemble learning algorithms, Random Forest and XGBoost regressors. Random Forest is a widely used algorithm that builds multiple decision trees and merges them to get a more accurate and stable prediction (Dutta et al., 2021). It is highly effective for regression tasks based on several features and is often used as a strong baseline model in sentiment analysis and rating prediction studies (Utami et al., 2024; Zhang et al., 2023). The second ensemble model is XGBoost, a highly efficient implementation of gradient boosting. It is considered as a strong model for regression and classification tasks with tabular data, by building decision trees sequentially where each new tree is trained to correct the errors of the previous ones (Chen & Guestrin, 2016). Its effectiveness in the domain of predicting customer ratings from topic-based features has been successfully demonstrated in recent studies (Yang et al., 2024).

Support Vector Regression (SVR) is included to represent a different family of algorithms from tree-based models. It works by finding an optimal solution by a loss function that enables the deviations from true values remain within a specified threshold which is called epsilon (Rozinajov´a et al., 2018). Previously, SVR is applied in product recommendations based on customer reviews (Yesodha et al., 2018) and predicting the sentiment of tourist reviews (Puh & Babac, 2023). K-Nearest Neighbors (KNN) are also included to serve as a simple but effective method that makes predictions based on the "k" most similar examples called neighbors in the training data. While it is mostly used in recommendation systems to find alternative products for customers (Anwar et al., 2022; Bahrani et al., 2024), it has also been used as a common benchmark in text classification and prediction tasks (Roy et al., 2018; Syeds & Thirupathy, 2023).

Finally, to represent the neural network approach Multilayer Perceptron (MLP) is applied. MLP is a type of artificial neural network that works by processing information through input, hidden and output neurons which allows it to model complex relationships between features (Chan et al., 2023). The application of various neural network architectures for the task of predicting ratings from review data is a widely adopted in the literature (Seo et al., 2017). By including MLP, the study can test whether a neural network model offers a performance advantage over tree-based or other classical models for the constructed feature sets.

3.7 Model Evaluation

To assess and understand the performance of the predictive models, a two-step process is applied. First the predictive accuracy of the models is evaluated using standard metrics. Second, an advanced interpretation method is used to explain the behavior of the best-performing models.

To measure the performance of the predictive models two common metrics, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are used. MAE calculates the average absolute difference between the predicted star ratings and the actual star ratings given by the customers. RMSE calculates the square root of the average of the squared differences between predicted and actual values. For both RMSE and MAE, a lower value indicates a more accurate model with better predictive performance.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (2)$$

3.8 Interpretations with Explainable AI

While metrics like RMSE and MAE can tell us how accurate the models are, they do not explain why models makes a specific prediction. The models such as XGBoost, Random Forest or MLP, are considered "black box" models since they are very complex and their internal logic is challenging for a human to understand (Hassija et al., 2024; Rai, 2020). To gain deeper insights and understand which factors are more influential in the predictive model, Shapley Additive Explanations (SHAP) proposed by Lundberg & Lee (2017) are employed in the framework. SHAP was selected over feature importance scores of machine learning models for several reasons. Unlike feature importance scores provided by tree-based models, SHAP values are grounded in game theory and provide a consistent and theoretically sound measure of each feature's contribution to individual predictions (Lundberg & Lee, 2017). Furthermore, SHAP has been increasingly adopted in hospitality and marketing research for interpreting machine learning models due to its ability to capture both the direction and magnitude of each feature's effect on the predicted outcome (Antolini et al., 2025; Wang et al., 2024).

4. Results

4.1 Dataset

In total, a corpus of 3976 reviews were collected within the time span between December 9 2007 and October 21 2025. Text preprocessing was implemented to reduce noise in the data. After converting all to lowercase non-information elements such as URLs, HTML tags, social media mentions, hashtags, and emojis were removed. Stopwords starting with a standard list of common English words, such as "the," "a," and "is" were also removed. To this, a custom list including the names of the restaurants, specific locations like "Istanbul" or "Bodrum," and other frequent but non-informative terms like "restaurant" was also added. By removing this combined list of words, it is ensured that the topic model focus on the content of the customer's feedback rather than on common or repetitive terms. Finally, duplicates and any reviews that were left with only one word were removed from the dataset. This entire procedure resulted in a final corpus of 3935 reviews ready for the topic modeling and feature engineering stages.

4.2 Topic Extraction with BERTopic

After pre-processing, the next step is to extract the underlying topics from the review corpus. For this task, BERTopic is used and it created two separate topic models to capture features at different levels of detail, a document-based and a sentence-based model.

Finding the right number of topics and ensuring they are meaningful is a critical task in topic modeling. Therefore, a systematic hyperparameter tuning process for both approaches were conducted to construct the best topic models. This process involved testing different configurations of BERTopic's key parameters, namely the number of neighbors (n_neighbors) in the UMAP dimension reduction algorithm and the minimum cluster size (min_cluster_size) in the HDBSCAN clustering algorithm. The n_neighbors' parameter directly influences UMAP's dimensionality reduction by defining the size of the local neighborhood around each data point that the algorithm seeks to represent in the reduced low-dimensional output (McInnes et al., 2018). The min_cluster_size parameter controls the minimum size of the clusters formed by HDBSCAN (McInnes et al., 2017), and has a direct influence on the number of topics created.

To select the best model configuration, a mixed approach, integrating quantitative analysis with researcher judgment, is applied as suggested by the author of the BERTopic library, Grootendorst (2022). First, the topic coherence score, which measures how semantically related the words within each topic are, was calculated. However, a high coherence score does not always guarantee that the topics are useful for human interpretation (Hoyle et al., 2021; Meaney et al., 2023; Rahimi et al., 2024). Therefore, this was combined with a qualitative evaluation based on researcher judgment. The main goals of the judgment were to protect the granularity of the topics by ensuring they were specific and detailed, and to disallow repetition, making it clear that different topics were distinctly separate from one another. For the document-based topic model, the optimal configuration was found with n neighbors set to 15 and minimum cluster size set to 15. This configuration initially produced a set of topics. After careful researcher inspection, some of the very similar or fragmented topics were manually combined to improve the clarity and distinction of the topic set. This manual curation resulted in a final set of 12 document-level topics. For the sentence-based topic model, a different configuration was found to be optimal. Here, the best parameters were n neighbors set to 15 and min cluster size set to 75. A similar process of manual inspection and combination was performed on the initial output. The final topics were a more granular set of 22 sentence-level clusters. These two distinct sets of topics then served as the basis for creating the four feature sets as described in section 3.5.

The results of the topic extraction and tuning process are summarized in **Table 1**. This table presents the final set of 12 main categories derived from the document-based model and the 22 specific topics derived from the sentence-based model. As the table shows, the 12 topics from the document-level model represent broad, high-level themes that describe the general focus of a review. For example, this approach identified general categories like 'Dining Experience', 'Menu Options', and 'Complaints' for each of the approaches, representative words for the found topics are also visible in **Table 1**.

When compared with document-level topics, the 22 topics from the sentence-level model provide a much more detailed and specific view. These topics can be understood as sub-topics or the individual components that make up the main categories. For instance, the main category of *Dining Experience* is broken down into three more specific topics at the sentence level, the general 'Dining Experience' itself, 'Food Presentation', and 'Interior Atmosphere'. Similarly, the broad topic of *Menu Options* is represented by five distinct and granular topics in the sentence level model, 'Menu Options', 'Desserts', 'Drinks', 'Wine Pairing and Menus', and 'Appetizers'. The top words in the **Table 1** not only provide evidence for the meaning of each topic, but also highlight the difference in granularity between the two approaches. For example, the main category of "Menu options" in document-level approach is defined by general words like *italian, good, service, menu* and *view*. Some of the top words like service and view are also related to other found topics. In contrast, the topics found in the sentence-level approach are defined by much more precise keywords. "Menu options" topic in this approach have top words such as *course, menu, tasting* and all the top five words are directly related with the topic. This hierarchical structure between document-level and sentence-level approaches demonstrates an insightful picture, showing that the sentence-based approach

successfully decomposes the general themes of a review into specific components, which further provide details about user experience. This specific view provided by the sentence-level topics is valuable for the feature sets used in the predictive models.

Table 1. Final topic hierarchy with representative words.

Document-level topics (Top 5 Words)	Sentence-level topics	Sentence-level top 5 representative words
Dining Experience *(wine, food, menu, view, service)*	Dining Experience Food Presentation Interior Atmosphere	*food, experience, chef, service,* *good presentation, food, plate, plates,* *delicious atmosphere, music, decor, ambiance, nice*
Chef's team *(thank, service, experience, food, chef)*	Gratitude for Chef's Team	*thank, care, like thank, emir, bey*
Rooftop Bar Dining *(bar, view, food, views, rooftop)*	Rooftop Bar Dining	*bar, view, terrace, rooftop, city*
Michelin Star Dining *(michelin, star, michelin star, food, service)*	Michelin Star Dining Guidance from Reviews Reputation	*michelin, star, michelin star, starred reviews,* *recommend, highly, recommendation, review* *world, top, best, best world, list*
Menu Options *(italian, good, service, menu, view)*	Menu Options Desserts Drinks Wine Pairing and Menus Appetizers	*course, menu, courses, tasting, tasting menu* *dessert, cream, ice, desserts, ice cream* *cocktails, cocktail, drinks, food,* *bar wine, wines, menu, wine list, list* *amuse, bouche, amuse bouche, bread, bouches*
View *(view,food,service, great, dinner)*	View	*floor, view, top, top floor, bosphorus*
Service and Staff *(food, service, great, staff, amazing)*	Staff Service Interaction	*service, staff, friendly, waiter, attentive*
Cuisine Variety *(lamb, menu, good, food, course)*	Main Meat Dishes Portion Size	*lamb, fish, octopus, main, beef* *portions, small, portions small, small portions, portion*
Olive Garden *(olive, trees, olive trees, menu, chef)*	Olive Garden	*olive, trees, olive trees, garden, olive oil*
Value and Price *(food, good, expensive, service, view)*	Value and Price	*price, worth, expensive, prices, high*
Special Occasions *(birthday, great, food, service, celebrate)*	Special Occasions	*birthday, evening, special, night, celebrate*
Complaints *(food, view, price, disappointing, worst)*	Expectations Reservation Process	*expectations, disappointed, high, high expectations,* *really table, reservation, tables, advance, make, booking*

4.3 Hyperparameter Tuning for Machine Learning Models

The performance of machine learning models is highly dependent on their internal settings, which are called hyperparameters. To ensure that each of the five machine learning models performed at its best, a systematic hyperparameter tuning process was conducted for each of the four feature sets, independently. The optimal combination of parameters for each model in the rating prediction task was searched. **Table 2** provides a summary of the hyperparameters and the range of values tested for each of the five machine learning models.

For searching the best parameters, Grid Search with 5-fold cross-validation is used (Chang & Lin, 2011). This technique systematically builds and evaluates a model for every possible combination of the provided parameters. It uses a cross-validation process to find the combination that produces the most accurate and generalizable model. The parameter combination that resulted in the best cross-validation performance for each model was selected. The optimized model was used to generate final results for benchmark comparison. The best parameters obtained from each of the four feature sets are demonstrated in **Table 3**.

Table 2. Hyperparameter search grids for model tuning.

Model	Hyperparameter	Values tested
Random Forest	n_estimators	[100, 200, 300]
	max_depth	[10, 20, 30, None]
	min_samples_leaf	[1, 2, 4]
	min_samples_split	[2, 5, 10]
XGBoost	n_estimators	[100, 200, 400]
	learning_rate	[0.01, 0.1, 0.2]
	max_depth	[3, 5, 7]
	subsample	[0.7, 1.0]
	colsample_bytree	[0.7, 1.0]
Support Vector (SVR)	kernel	['linear', 'rbf']
	C (for linear)	[0.1, 1, 10, 100]
	C (for rbf)	[1, 10, 100, 1000]
	gamma (for rbf)	['scale', 'auto']
K-Nearest Neighbors (KNN)	n_neighbors	[3, 5, 7, 9, 11, 15]
	weights	['uniform', 'distance']
	metric	['euclidean', 'manhattan']
Multilayer Perceptron (MLP)	hidden_layer_sizes	[(50,), (100,), (50, 25)]*
	activation_function	['relu', 'tanh']
	alpha	[0.0001, 0.001, 0.01]
	max_iter	[500, 1000]

*Note: For hidden layer sizes, the notation (50,) represents one hidden layer with 50 neurons, while (50, 25) represents two hidden layers.

Table 3. Optimal hyperparameters found for each feature set.

Model	Hyperparameter	Document-based	Topic proportions	Average sentiments	Weighted sentiments
Random Forest	n_estimators	200	300	300	300
	max_depth	10	10	None	30
	min_samples_leaf	2	4	4	4
	min_samples_split	10	10	10	14
XGBoost	n_estimators	400	400	200	400
	learning_rate	0.01	0.01	0.1	0.1
	max_depth	3	3	3	3
	subsample	0.7	0.7	1.0	0.7
	colsample_bytree	1.0	0.7	0.7	0.7
SVR	kernel	'rbf'	'rbf'	'rbf'	'rbf' 1 'scale'
	C	100	1	1	
	gamma	'scale'	'scale'	'scale'	
KNN	n_neighbors	15	15	15	15
	weights	'uniform'	'uniform'	'uniform'	'uniform'
	metric	'euclidean'	'manhattan'	'manhattan'	'euclidean'
MLP	hidden_layer_sizes	(100,)	(50,)	(50,)	(50,)
	activation	'relu'	'relu'	'relu'	'relu'
	alpha	0.0001	0.0001	0.001	0.01
	max_iter	500	500	500	500
	solver	'adam'	'adam'	'adam'	'adam'

4.4 Comparative Performance of Prediction Models

After the hyperparameter tuning process, each of the five models was trained on the four different feature sets. For each set, 80% is used for training, and the remaining 20% was used for testing. There was a total of twenty models to be compared. The performance of all the models was evaluated on the test set using RMSE and MAE. The complete results are presented in **Table 4**.

The most significant finding from this comparison is the remarkable improvement in prediction accuracy when sentiment is included in the features. The first two feature sets, which are based on the presence of topics, show relatively higher errors, with RMSE values generally between 0.92 and 1.05. On the other hand, the feature sets that incorporate sentiment, namely *Average Sentiments* and *Weighted Sentiments*, achieve much lower errors, with RMSE values dropping to a range of approximately 0.67 to 0.79. This result clearly shows that understanding how a customer feels about a topic is much more predictive of their final rating than simply knowing what they talked about.

Furthermore, when comparing the two sentiment-based approaches, *Weighted Sentiments* feature set consistently produced the best results. For almost every model, this feature set, which combines both the emotional tone and the focus of the reviewer, achieved a lower RMSE and MAE than *Average Sentiments* feature set. For example, the Random Forest model's RMSE improved from 0.6932 to 0.6699 when using the weighted feature. This demonstrates the value of weighting the sentiment with topic prominence, as it more accurately captures the reviewer's orientation.

Table 4. Comparative performance of prediction models (RMSE and MAE).

Feature group	Model	RMSE	MAE
Features Based on Topic Presence	*Document based*		
	Random Forest	0.9494	0.6718
	XGBoost	0.9356	0.6834
	Support Vector	1.0207	0.5968
	KNN	0.9644	0.6787
	Multilayer Perceptron	0.9295	0.6682
	Topic Props (sentence based)		
	Random Forest	0.9685	0.7159
	XGBoost	0.9679	0.7294
	Support Vector	1.0533	0.6307
	KNN	0.9943	0.7180
	Multilayer Perceptron	1.0517	0.7741
Features Incorporating Sentiment	*Average Sentiments (sentence based)*		
	Random Forest	0.6932	0.4781
	XGBoost	0.7035	0.5083
	Support Vector	0.7648	0.4998
	KNN	0.7465	0.4980
	Multilayer Perceptron	0.7949	0.5678
	Weighted Sentiments (sentence based)		
	Random Forest	**0.6699**	0.4571
	XGBoost	**0.6684**	0.4657
	Support Vector	0.6915	**0.4415**
	KNN	0.7342	0.5196
	Multilayer Perceptron	0.6768	0.4747

Finally, the best performing model overall was the XGBoost model trained on the Weighted Sentiments feature set. This model achieved the lowest error, with a RMSE of 0.6684 and a MAE of 0.4657. It is important to note that the Random Forest (RMSE = 0.6699) and MLP (RMSE = 0.6768) models also performed well with this feature set, confirming its general effectiveness across different model architectures.

4.5 Interpretation of the Models Using SHAP

To understand how the models, make their prediction, more specifically, which features have more influence on the rating prediction, SHAP framework is used. In the first part, the focus was on interpreting the models that were trained on features representing only topic presence. By analyzing these models, the study can

see how the simple act of focusing on a topic influences the predicted rating. Then in the second part, the focus was on interpreting sentiment-based features to understand how incorporating sentiment in the models prevail the rating prediction.

4.5.1 Insights from Topic Prevalence

To understand the predictive power of the features only represent the focus of a topic, the models that were trained on topic prevalence alone were analyzed. These are document-based and sentence-based topic prominence feature sets. The results of SHAP for the best performing models for these features, namely XGBoost, are represented in **Figures 2(a)** and **2(b)**. Both figures, reveal a consistent story about how a reviewer's focus can predict their rating.

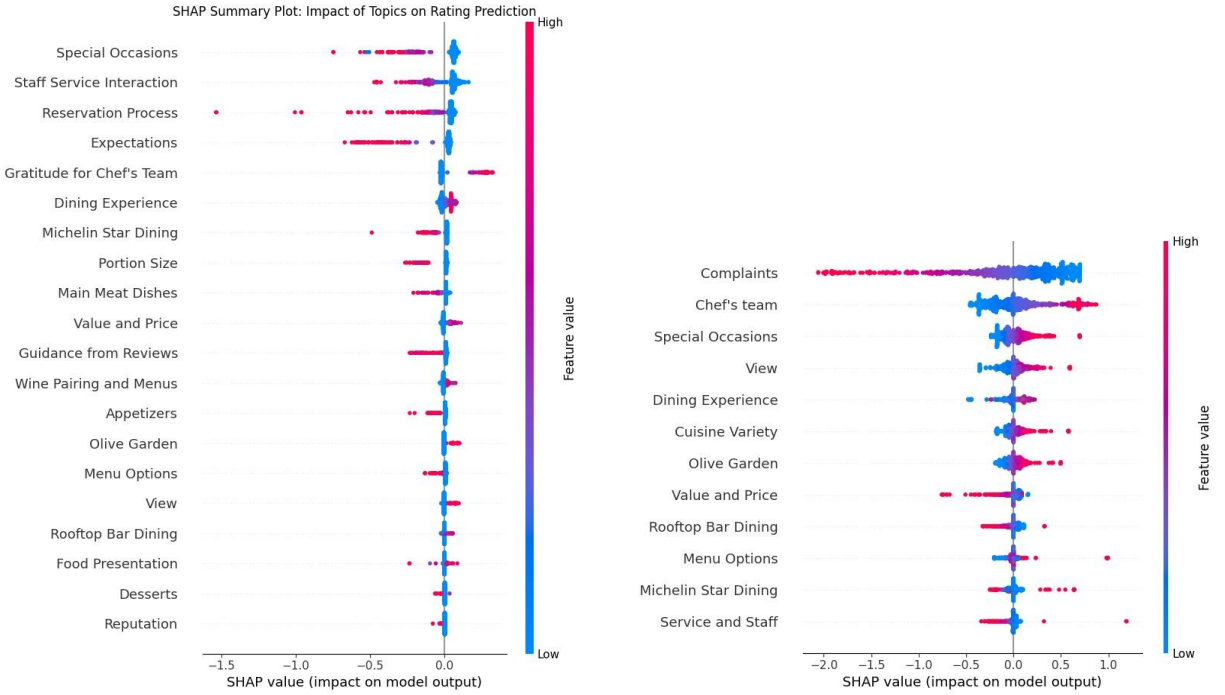
In SHAP diagrams, red dots represent the increasing values for the feature while blue dots represent lower, decreasing values. When the SHAP plot for the document-level model in **Figure 2(b)** is examined, the most influential topic was *Complaints*. When the probability of a review being about complaints is high, the impact on the rating is strongly negative. This finding is logical since if the main point of a review is to complain, it is expected to have a low star rating. Conversely, the most important topics were *Chef's Team* and *Special Occasions*. For both of these a high probability has a strong positive impact on the rating. This shows, if a review focus on the chef or celebrating an event, it is a clear signal of a very positive experience.

The SHAP plot for the sentence-level as depicted in **Figure 2(a)**, provides a more detailed but consistent view. Similar to the document-level plot, the strongest positive drivers were *Special Occasions*, *Gratitude for Chef's Team*, *Reservation Process* and *Expectations*. This granular view breaks down the general Complaints category into its specific components. These granular topics are strong negative drivers of customer ratings. For all of these, a high proportion of text dedicated to these topics, has a strong negative impact on ratings. This finding indicates that when customers write extensively about the service or reservation process, it means they have had a negative experience regarding these features, which is a powerful signal of a low rating.

4.5.2 Insights from Sentiment-based Features

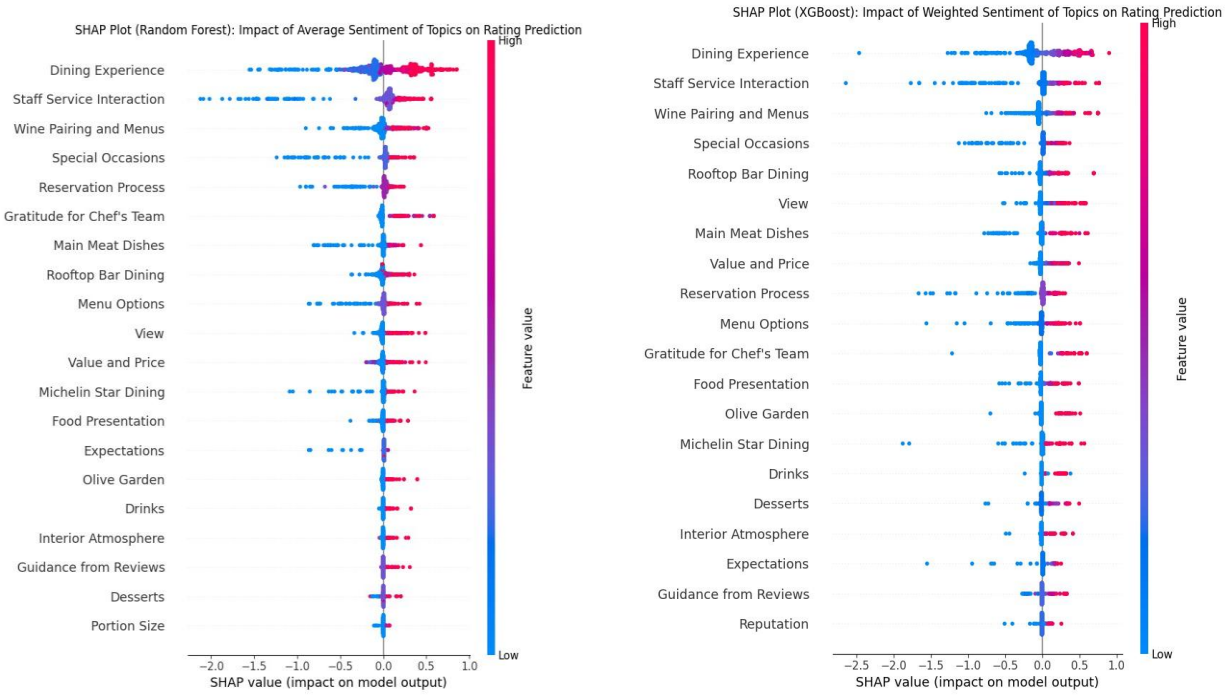
As shown in our performance evaluation, the models trained on sentiment-based features achieved a remarkable improvement in accuracy. To understand the reasons for this success, SHAP was used to interpret the predictions of the best models. SHAP plots are obtained from the overall best model (XGBoost with Weighted Sentiments) and compared with the best performing model for Average sentiments (Random Forest). The SHAP summary plots for these two models are shown in **Figures 3(a)** and **3(b)**.

The most important finding is the noticeable consistency between the two models. Both plots (**Figures 3(a)** and **3(b)**) show that the same topics are the most influential and they affect the rating prediction in the same way. This gives high confidence that the findings are robust and accurately reflect the drivers of customer ratings. Both models agree that Dining Experience and Staff Service Interaction are the top most influential topics. For both, there is a powerful two-way relationship meaning high positive sentiment (red dots) strongly increases the predicted rating, while low negative sentiment (blue dots) strongly decreases it. This finding shows that overall experience and the quality of human service are the fundamental pillars of a high or low rating.



(a) Impact of sentence-level topics (b) Impact of document-level topics

Figure 2. SHAP summary plots for models trained on topic prevalence features.



(a) Impact of average sentiments (b) Impact of weighted sentiments

Figure 3. SHAP summary plots for models trained on sentiment-based features.

Another important finding based on **Figures 3(a)** and **3(b)** is the asymmetric pattern in the influence of *Staff Service interaction*, *Special occasions* and *Reservation process*. For all of these topics, strong negative sentiment has a strong negative impact on the rating. However, positive sentiment provides almost no benefit, with the points clustered very close to zero. This means the negative impact of an unpleasant experience regarding these factors is often more severe than the positive impact of a pleasant experience. A failure is punished by the customers, but a success is not rewarded that much, probably because they are basic unspoken expectations. This finding supports the SHAP results of topic prevalence alone, which found these factors as important drivers of low rating.

5. Discussion

This study developed and tested a framework for predicting customer ratings of Michelin-starred restaurants using topic modelling and machine learning models. The results showed that the feature sets that included sentiment were more accurate than those without. The feature set in which sentiment is weighted with topic prevalence, as shown in (Birim & Kazancoglu, 2026), produced the best predictive performance. Explainable AI analysis further revealed that Dining Experience and Chef's Team were the main drivers of high ratings, while Staff Service Interactions, Special Occasions, and the Reservation process were the key factors associated with lower ratings. Specifically, Special Occasions exhibits an asymmetric pattern. Negative experiences related to special occasions strongly reduce the predicted rating, while positive experiences provide near-zero increase, suggesting that customers expect these moments to be handled perfectly and penalize failures more than they reward successes.

These findings are consistent with and build upon the existing literature in this field. The conclusion that sentiment is a powerful predictor of star ratings aligns with many previous studies that have established a strong link between the emotional tone of a review and the evaluation score of customers (Pang & Lee, 2008). Furthermore, the SHAP analysis identified specific topics that are well-known drivers of customer satisfaction. For example, the finding that *Dining Experience* and *Staff Service Interaction* are the most powerful predictors is supported by research that identifies service quality and the overall atmosphere as fundamental components of the luxury dining experience (Namkung & Jang, 2008; Tsaur & Lo, 2020). The one-sided negative impact of topics like Staff Service Interaction and Reservation Process also supports that certain elements are expected to be perfect and only draw attention when they fail (Sulek & Hensley, 2004). In Michelin-starred restaurants, the dining experience extends beyond food quality to encompass service, atmosphere, and presentation. These findings highlight the critical need for staff training focused not only on competence, but on creating positive and memorable interactions.

In elegant restaurants like Michelin-starred, customers generally have high expectations due to assurance of quality, and when their expectations are not met, they exhibit negative emotions (Rita et al., 2023). The strong influence of *Special Occasions* found in this study also supports existing research about delivering authentic services to achieve customer delight, which shows that exceptional and memorable events are key drivers of the high satisfaction level (Kim et al., 2024). The SHAP analysis provides several clear, actionable insights for restaurant managers. The strong impact of the *Special Occasions* topic shows a clear opportunity for managers. By actively identifying and acknowledging guests who are celebrating birthdays or anniversaries, and by making a small extra effort to enhance their celebration, restaurants can create highly memorable experiences that are very likely to result in higher ratings and positive reviews. Personalized gestures such as special menus or chef introductions can transform these occasions into memorable experiences that strengthen customer loyalty (Kim et al., 2025). The asymmetric impact of *Reservation Process* provides a warning for the managers. The reservation process is the first point of contact in the customer experience. Effective management of this process will lead to operational and emotional improvements. Customers often wait months to get a table at Michelin-starred restaurants, so the

reservation process can be long. A difficult or flawed booking system is a major driver of low customer satisfaction and low ratings. Therefore, managers should prioritize investing in making the reservation system well-working with the least number of flaws, such as digital waiting lists can make the reservation process more transparent. These improvements will facilitate a smoother reservation process. Restaurants using AI-powered reservation systems based on digital platforms can enhance the overall customer experience, streamline operations, and increase efficiency. Customer data collected during the reservation process (special occasions, preferred dishes and wines, preferred table, etc.) will facilitate the provision of personalized service by CRM-integrated reservation software (Samanta et al., 2026). These developments and improvements reduce dissatisfaction and offer a better dining experience.

In contrast, the strong positive influence of *Gratitude for Chef's Team* indicates that creating opportunities for customers to feel a connection to the culinary staff, such as through open kitchens or communicating with the chef at the table, could be a powerful strategy for generating exceptionally high ratings. Chefs play a key role in fostering innovation and delivering memorable culinary experiences (Mrusek et al., 2021; Ottenbacher & Harrington, 2007).

6. Conclusion

This study proposes a framework that includes key factors influencing customer satisfaction based on customer reviews at Michelin-starred restaurants. In order to accomplish this, BERTopic, a transformer-based topic model, was used to extract factor topics from a dataset of reviews from Michelin-starred restaurants. Accordingly, by interpreting the best-performing model, the factors that best predict customer satisfaction were identified. According to the SHAP analysis results, feature sets that include sentiment analysis performed better than those based solely on theme presence. As shown in Birim and Kazancoglu, 2025, the feature set where sentiment analysis was weighted by theme prevalence produced the best prediction performance, while the XGBoost model achieved the lowest error. According to the SHAP analysis results, different types of factors such as dining experience and chef's team were identified as significant performance factors in the models. Staff service interactions, special occasions, and the reservation process were identified as factors that decreased the rating when customer had a negative feeling. Therefore, the study findings go beyond prediction and offer actionable insights.

This study provides actionable insights into the key drivers of customer satisfaction. It is important to note that there are some limitations which in turn opens several future research avenues. The dataset was focused exclusively on one-starred Michelin Restaurants in Turkey, and therefore, the findings may not be directly generalizable to restaurants in different cultural contexts or at different quality tiers of Michelin (e.g. two-star, Green Star or non-Michelin restaurants). Additionally, the analysis relies on a single, pre-trained sentiment analysis model. Although the chosen model is fine-tuned for reviews, different models can have their own inherent biases and may produce slightly different sentiment scores for the same sentence. Future research could test the robustness of these findings by comparing the results using several different sentiment analysis tools. Future studies should apply this same analytical framework to a more diverse, international dataset to conduct a comparative analysis of satisfaction drivers across different countries and restaurant types. Furthermore, future work could employ native language NLP models to avoid the translation step and to obtain language-specific insights. Such research would help to build a more comprehensive and globally applicable framework to understand the key drivers of customer satisfaction in the dining industry.

Conflicts of Interest
The author(s) have no conflicting interests that are relevant to the content of this article.

Acknowledgments
The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

AI Disclosure
The author(s) confirm that generative AI tools were not used during the preparation of this article.

References

Amelia, M., & Garg, A. (2016). The first impression in a fine-dining restaurant. A study of C restaurant in Tampere, Finland. *European Journal of Tourism, Hospitality, and Recreation*, *7*(2), 100-111. https://doi.org/10.1515/ejthr-2016-0012

Amirifar, T., Lahmiri, S., & Zanjani, M.K. (2023). An NLP-deep learning approach for product rating prediction based on online reviews and product features. *IEEE Transactions on Computational Social Systems*, *11*(6), 8156-8168. https://doi.org/10.1109/TCSS.2023.3290558

Antolini, F., Cesarini, S., & Terraglia, I. (2025). Explaining tourism expenditure patterns in Italy: a comparative study of domestic and ıncoming tourists using XGBoost and SHAP. *Social Indicators Research*, *180*(1), 369-382. https://doi.org/10.1007/s11205-025-03685-9

Anwar, T., Uma, V., Hussain, M.I., & Pantula, M. (2022). Collaborative filtering and kNN based recommendation to overcome cold start and sparsity issues: a comparative analysis. *Multimedia Tools and Applications*, *81*(25), 35693-35711. https://doi.org/10.1007/s11042-021-11883-z

Aparicio-Aparicio, I., Ruescas-Nicolau, M.-A., Duen˜as, L., S´anchez-Jim´enez, J.L., S´anchez-S´anchez, M.L., & Alc´antara, E. (2024). Discovering the relationship between attributes of facial masks and review rating in online customer reviews using explainable artificial intelligence (xai). *Applied Artificial Intelligence*, *38*(1), 2411780. https://doi.org/10.1080/08839514.2024.2411780

Bahrani, P., Minaei-Bidgoli, B., Parvin, H., Mirzarezaee, M., & Keshavarz, A. (2024). A new improved KNN-based recommender system. *The Journal of Supercomputing*, *80*(1), 800-834. https://doi.org/10.1007/s11227-023-05447-1

Bang, D., Choi, K., & Kim, A.J. (2022). Does Michelin effect exist? An empirical study on the effects of Michelin stars. *International Journal of Contemporary Hospitality Management*, *34*(6), 2298-2319. https://doi.org/10.1108/IJCHM-08-2021-1025

Barrera-Barrera, R. (2023). Identifying the attributes of consumer experience in Michelin-starred restaurants: a text-mining analysis of online customer reviews. *British Food Journal*, *125*(13), 579-598. https://doi.org/10.1108/BFJ-05-2023-0408

Berry, L.L., Wall, E.A., & Carbone, L.P. (2006). Service clues and customer assessment of the service experience: lessons from marketing. *Academy of Management Perspectives*, *20*(2), 43-57. https://doi.org/10.5465/amp.2006.20592105

Bertan, S. (2016). The Evaluation of michelin star restaurants. *Journal of Human Sciences*, *13*(2), 3221-3230. https://doi.org/10.14687/jhs.v13i2.3807

Birim, S.O., & Kazancoglu, I. (2026). Decoding customer satisfaction in Michelin green star restaurants: BERTopic modelling and sentiment analysis. *British Food Journal*, *128*(4), 1274-1313. https://doi.org/10.1108/BFJ-07-2025-0883

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993-1022.

Castillo-Manzano, J.I., Castro-Nuno, M., Lopez-Valpuesta, L., & Zarzoso, A. (2021). Quality versus quantity: an assessment of the impact of Michelin-starred restaurants on tourism in Spain. *Tourism Economics*, *27*(5), 1166-1174. https://doi.org/10.1177/1354816620914838

Cha, T., & Lee, D. (2024). Sentencelda: discriminative and robust document representation with sentence level topic model. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics* (Vol. 1, pp. 521-538). Association for Computational Linguistics. St. Julian's, Malta. https://doi.org/10.18653/v1/2024.eacl-long.31

Chan, K.Y., Abu-Salih, B., Qaddoura, R., Al-Zoubi, A.M., Palade, V., Pham, D.S., Ser, J.D., & Muhammad, K. (2023). Deep neural networks in the cloud: review, applications, challenges and research directions. *Neurocomputing*, *545*, 126327. https://doi.org/10.1016/j.neucom.2023.126327

Chang, C.C., & Lin, C.J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on İntelligent Systems and Technology*, *2*(3), 1-27. https://doi.org/10.1145/1961189.1961199

Chang, Y.Y., Cheng, C.C., Tsai, M.C., & Xie, P.W. (2025). Bridging the knowledge gap of memorable dining experience at Michelin-starred restaurants: insights from a new two-dimensional strategic matrix. *British Food Journal*, *127*(6), 2035-2064. https://doi.org/10.1108/BFJ-12-2023-1066

Chaturvedi, P., Kulshreshtha, K., Tripathi, V., & Agnihotri, D. (2024). Investigating the impact of restaurants' sustainable practices on consumers' satisfaction and revisit intentions: a study on leading green restaurants. *Asia-Pacific Journal of Business Administration*, *16*(1), 41-62. https://doi.org/10.1108/APJBA-02-2023-0077

Chen, C.M., & Lin, Y.C. (2025). The non-linear price effect on guest satisfaction of the Michelin-starred restaurants. *Current Issues in Tourism*, *28*(24), 3903-3907. https://doi.org/10.1080/13683500.2024.2425987

Chen, T., & Guestrin, C. (2016). Xgboost: a scalable tree boosting system. In *Proceedings of the 22nd Acm Sigkdd İnternational Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM. New York, USA. https://doi.org/10.1145/2939672.2939785

Chiang, C.F., & Guo, H.W. (2021). Consumer perceptions of the Michelin guide and attitudes toward Michelin-starred restaurants. *International Journal of Hospitality Management*, *93*, 102793. https://doi.org/10.1016/j.ijhm.2020.102793

Choi, J., Park, J., Jeon, H., & Asperin, A.E. (2021). Exploring local food consumption in restaurants through the lens of locavorism. *Journal of Hospitality Marketing & Management*, *30*(8), 982-1004. https://doi.org/10.1080/19368623.2021.1923608

Daries, N., Marine-Roig, E., Ferrer-Rosell, B., & Cristobal-Fransi, E. (2021). Do high-quality restaurants act as pull factors to a tourist destination?. *Tourism Analysis*, *26*(2-3), 195-210. https://doi.org/10.3727/108354221X16104618764020

de Lima, B.C., Baracho, R.M.A., Mandl, T., & Porto, P.B. (2023). Reactions to science communication: discovering social network topics using word embeddings and semantic knowledge. *Social Network Analysis and Mining*, *13*(1), 119. https://doi.org/10.1007/s13278-023-01125-5

Dutta, P., Paul, S., & Kumar, A. (2021). Comparative analysis of various supervised machine learning techniques for diagnosis of COVID-19. In *Electronic Devices, Circuits, and Systems for Biomedical Applications*: *Challenges and Intelligent Approach* (pp. 521-540). Academic Press. USA. https://doi.org/10.1016/B978-0-323-85172-5.00020-4

Farrugia, M. (2024). *How Michelin guide recommendations are affecting the restaurant industry in Malta*? [Bachelor's dissertation, Institute of Tourism Studies, Malta]. The Online Open Access Repository of the Institute of Tourism Studies (Malta).

Gallinucci, E., Golfarelli, M., & Rizzi, S. (2015). Advanced topic modeling for social business intelligence. *Information Systems*, *53*, 87-106. https://doi.org/10.1016/j.is.2015.04.004

Grootendorst, M. (2022). Bertopic: neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*. https://doi.org/10.48550/arXiv.2203.05794

Grootendorst, M. (2024). Topic Distributions — BERTopic Documentation [Accessed: 2024-05-16]. https://maartengr.github.io/BERTopic/getting_started/distribution/distribution.html

Harrington, R.J., Fauser, S.G., Ottenbacher, M.C., & Kruse, A. (2013). Key information sources impacting Michelin restaurant choice. *Journal of Foodservice Business Research*, *16*(3), 219-234. https://doi.org/10.1080/15378020.2013.810534

Hassija, V., Chamola, V., Mahapatra, A., Sahil, Wang, F.Z., & Singh, N. (2024). Interpreting blackbox models: a review on explainable artificial intelligence. *Cognitive Computation*, *16*(1), 45-74. https://doi.org/10.1007/s12559-023-10179-8

Ho, C., Tu, H.S., Anh, N.N., Tuan, P.M., Anh, N.T.N., & Ha, P.T.M. (2021). Factors affecting customer satisfaction and repurchase intention: a study of full-service trendy restaurants in Vietnam. *Procedia Business and Financial Technology*, *1*(1), 75-180.

Hoyle, A., Goel, P., Hian-Cheong, A., Peskov, D., Boyd-Graber, J., & Resnik, P. (2021). Is automated topic model evaluation broken? the incoherence of coherence. *Advances in Neural İnformation Processing Systems*, *34*, 2018-2033.

Huang, F., Yuan, C., Bi, Y., Lu, J., Lu, L., & Wang, X. (2022). Multi-granular document-level sentiment topic analysis for online reviews. *Applied Intelligence*, *52*(7), 7723-7733. https://doi.org/10.1007/s10489-021-02817-1

Jiang, H., Zhou, R., Zhang, L., Wang, H., & Zhang, Y. (2019). Sentence level topic models for associated topics extraction. *World Wide Web*, *22*, 2545-2560. https://doi.org/10.1007/s11280-018-0639-1

Kim, J.Y., & Hwang, J. (2022). Who is an evangelist? food tourists' positive and negative e-WOM behavior. *International Journal of Contemporary Hospitality Management*, *34*(2), 555-577. https://doi.org/10.1108/IJCHM-05-2021-0610

Kim, S., Kim, M., & Choi, L. (2024). "Going the extra mile": an integrative model of customer delight. *International Journal of Contemporary Hospitality Management*, *36*(4), 1193-1212. https://doi.org/10.1108/IJCHM-06-2023-0857

Kim, S., Riswanto, A.L., Sherpa, S., & Kim, H.S. (2025). Generation Z's Michelin choices: survey and big data insights into Thailand's Michelin-starred restaurants. *International Journal of Tourism Research*, *27*(5), e70135. https://doi.org/10.1002/jtr.2711

Lee, J.-S., & Hwang, J. (2011). Luxury marketing: the influences of psychological and demographic characteristics on attitudes toward luxury restaurants. *International Journal of Hospitality Management*, *30*(3), 658-669. https://doi.org/10.1016/j.ijhm.2010.12.008

Liu, C., Cai, X., & Zhu, H. (2015). Eating out ethically: an analysis of the influence of ethical food consumption in a vegetarian restaurant in Guangzhou, China. *Geographical Review*, *105*(4), 551-565. https://doi.org/10.1111/j.1931-0846.2015.12097.x

Lu, W., & Stepchenkova, S. (2015). User-generated content as a research mode in tourism and hospitality applications: topics, methods, and software. *Journal of Hospitality Marketing & Management*, *24*(2), 119-154. https://doi.org/10.1080/19368623.2014.907758

Lundberg, S.M., & Lee, S.I. (2017). A unified approach to interpreting model predictions. *Advances in Neural İnformation Processing Systems*, *30*, 4768-4777.

Luo, Y., & Xu, X. (2021). Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic. *International Journal of Hospitality Management*, *94*, 102849. https://doi.org/10.1016/j.ijhm.2021.102849

Martins, H.J. (2022). *Michelin starred restaurants as drivers of gastronomic tourism and the main experience dimensions - the Portuguese case* [Master's thesis, ISCTE - Instituto Universit´ario de Lisboa]. http://hdl.handle.net/10071/27891

McInnes, L., Healy, J., & Astels, S. (2017). Hdbscan: hierarchical density-based clustering. *The Journal of Open Source Software*, *2*(11), 205. https://doi.org/10.21105/joss.00205

McInnes, L., Healy, J., & Melville, J. (2018). Umap: uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*. https://arxiv.org/abs/1802.03426

Meaney, C., Stukel, T.A., Austin, P.C., & Schull, M.J. (2023). Quality indices for topic model selection and evaluation: A literature review and case study. *BMC Medical Informatics and Decision Making*, *23*(1), 132. https://doi.org/10.1186/s12911-023-02216-1

Meng, Y., Yang, N., Qian, Z., & Zhang, G. (2021). What makes an online review more helpful: an interpretation framework using XGBOOST and SHAP values. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(3), 466-490. https://doi.org/10.3390/jtaer16030029

Miah, M.S.U., Kabir, M.M., Sarwar, T.B., Safran, M., Alfarhood, S., & Mridha, M.F. (2024). A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM. *Scientific Reports*, *14*(1), 9603. https://doi.org/10.1038/s41598-024-60210-7

Michelin Guide (2022). What is a Michelin green star? [Accessed: 2025-03-18]. https://guide.michelin.com/qa/en/article/features/what-is-a-michelin-green-star-dubai

Mrusek, N., Ottenbacher, M.C., & Harrington, R.J. (2021). The impact of sustainability and leadership on the innovation management of Michelin-starred chefs. *Sustainability*, *14*(1), 330. https://doi.org/10.3390/su14010330

Namkung, Y., & Jang, S. (2008). Are highly satisfied restaurant customers really different? a quality perception perspective. *International Journal of Contemporary Hospitality Management*, *20*(2), 142-155. https://doi.org/10.1108/09596110810852131

Nguyen, V.-H., & Ho, T. (2023). Analysing online customer experience in hotel sector using dynamic topic modelling and net promoter score. *Journal of Hospitality and Tourism Technology*, *14*(2), 258-277. https://doi.org/10.1108/JHTT-04-2021-0116

NLP Town (2024). bert-base-multilingual-uncased-sentiment (Revision 2a0c1d) [Pre-trained language model]. Hugging Face. https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment

Ottenbacher, M., & Harrington, R.J. (2007). The innovation development process of Michelin-starred chefs. *International Journal of Contemporary Hospitality Management*, *19*(6), 444-460. https://doi.org/10.1108/09596110710775110

Pacheco, L. (2018). An analysis of online reviews of upscale Iberian restaurants. *Dos Algarves: A Multidisciplinary e-Journal*, *32*, 38-53. https://doi.org/10.18089/damej.2018.32.3

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1), 1-135. https://doi.org/10.1561/1500000011

Pantelidis, I.S. (2010). Electronic meal experience: a content analysis of online restaurant comments. *Cornell Hospitality Quarterly*, *51*(4), 483-491. https://doi.org/10.1177/1938965510378574

Park, E., Chae, B., & Kwon, J. (2020). The structural topic model for online review analysis: comparison between green and non-green restaurants. *Journal of Hospitality and Tourism Technology*, *11*(1), 1-17. https://doi.org/10.1108/JHTT-08-2017-0075

Puh, K., & Babac, M.B. (2023). Predicting sentiment and rating of tourist reviews using machine learning. *Journal of Hospitality and Tourism Insights*, *6*, 1188-1204. https://doi.org/10.1108/JHTI-022022-0078

Rahimi, H., Mimno, D., Vigly, J.H., Naacke, H., Constantin, C., & Amann, B. (2024). Contextualized topic coherence metrics. In *Findings of the Association for Computational Linguistics: EACL 2024* (pp. 1760-1773). Association for Computational Linguistics, St. Julian's, Malta.

Rai, A. (2020). Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137-141. https://doi.org/10.1007/s11747-019-00710-5

Reis, C. (2024). Online reviews at the table: the influence of online reviews on Michelin-starred restaurants. Supervised by Prof. Helena Rodrigues, Master's thesis, Strategic Marketing, Universidade Católica Portuguesa (Portugal).

Rita, P., Vong, C., Pinheiro, F., & Mimoso, J. (2023). A sentiment analysis of Michelin-starred restaurants. *European Journal of Management and Business Economics*, *32*(3), 276-295. https://doi.org/10.1108/EJMBE-11-2021-0295

Roy, S.S., Kaul, D., Roy, R., Barna, C., Mehta, S., Misra, A. (2018). Prediction of customer satisfaction using naive bayes, multiclass classifier, K-star and IBK. In: Balas, V., Jain, L., Balas, M. (eds) *Soft Computing Applications. SOFA 2016* (Vol. 634). Springer, Cham. Arad, Romania. https://doi.org/10.1007/978-3-319-62524-9_12

Rozinajová, V., Ezzeddine, A.B., Lóderer, M., Loebl, J., Magyar, R., & Vrablecová, P. (2018). Computational intelligence in smart grid environment. In: Sangaiah, A.K., Sheng, M., & Zhang, Z. (eds) *Computational İntelligence for Multimedia Big Data on the Cloud with Engineering Applications* (pp. 23-59). Academic Press. MA, USA. https://doi.org/10.1016/B978-0-12-813314-9.00002-5

Sahoo, A., Chanda, R., Das, N., & Sadhukhan, B. (2023). Comparative analysis of bert models for sentiment analysis on twitter data. In *2023 9th International Conference on Smart Computing and Communications* (pp. 658-663). IEEE. India. https://doi.org/10.1109/ICSCC59169.2023.10335061

Samanta, R.N., Mathad, V.P., & Kamath, R. (2026). Reception automation and guest experience: balancing efficiency and personalization in hotel front offices. In *Robotics in Hotel Services: Housekeeping, Reception, and Concierge Services* (pp. 121-152). IGI Global Scientific Publishing. USA.

Sanchez-Franco, M.J., & Rey-Moreno, M. (2022). Do travelers' reviews depend on the destination? an analysis in coastal and urban peer-to-peer lodgings. *Psychology and Marketing*, *39*, 441-459. https://doi.org/10.1002/mar.21608

Sangkaew, N., Nanthaamornphong, A., & Phucharoen, C. (2025). Understanding tourists' perception toward local gourmet consumption in the creative city of gastronomy: Factors influencing consumer satisfaction and behavioral intentions. *Journal of Quality Assurance in Hospitality & Tourism*, *26*(2), 332-359. https://doi.org/10.1080/1528008X.2023.2268819

Schuckert, M., Peters, M., & Pilz, G. (2018). The co-creation of host–guest relationships via couchsurfing: a qualitative study. *Tourism Recreation Research*, *43*(2), 220-234. https://doi.org/10.1080/02508281.2017.1384127

Seo, S., Huang, J., Yang, H., & Liu, Y. (2017). Interpretable convolutional neural networks with dual local and global attention for review rating prediction. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (pp. 297-305). ACM, Como, Italy. https://doi.org/10.1145/3109859.3109890

Shahhosseini, M., & Nasr, A.K. (2024). What attributes affect customer satisfaction in green restaurants? An aspect-based sentiment analysis approach. *Journal of Travel and Tourism Marketing*, *41*, 472-490. https://doi.org/10.1080/10548408.2024.2306358

Sulek, J.M., & Hensley, R.L. (2004). The relative importance of food, atmosphere, and fairness of wait: the case of a full-service restaurant. *Cornell Hotel and Restaurant Administration Quarterly*, *45*(3), 235-247. https://doi.org/10.1177/0010880404265345

Svejenova, S., Mazza, C., & Planellas, M. (2007). Cooking up change in haute cuisine: Ferran adri`a as an institutional entrepreneur. *Journal of Organizational Behavior*, *28*(5), 539-561. https://doi.org/10.1002/job.461

Syeds, S., & Thirupathy, P. (2023). A novel approach for precision and recall estimation for star rating online customers based on positive movie reviews using naive bayes algorithm over K-nearest neighbour algorithm. In *International Conference on Environmental Engineering and Material Sciences: ICEEMS-2021* (Vol. 2655, No. 1, p. 020089). AIP Publishing LLC. India. https://doi.org/10.1063/5.0134731

Temiz, S., Budak, I., Kurto˘glu, R., & Atsiz, O. (2025). Understanding the luxury gastronomy experience through Michelin restaurants: a Netnography approach with LDA and sentiment analysis. *Journal of Foodservice Business Research*, 1-37. https://doi.org/10.1080/15378020.2024.2346740

Tirunillai, S., & Tellis, G.J. (2014). Mining marketing meaning from online chatter: strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research*, *51*(4), 463-479. https://doi.org/10.1509/jmr.12.0106

Tsao, W.C., & Hsieh, M.T. (2015). E-WOM persuasiveness: Do e-WOM platforms and product type matter? *Electronic Commerce Research*, *15*(4), 509-541. https://doi.org/10.1007/s10660-015-9193-9

Tsaur, S.-H., & Lo, P.-C. (2020). Measuring memorable dining experiences and related emotions in fine dining restaurants. *Journal of Hospitality Marketing & Management*, *29*(8), 887-910. https://doi.org/10.1080/19368623.2020.1718118

Utami, A., Permadi, D.F.H., Rosita, Y.D., & Unjung, J. (2024). Performance comparison of random forest (rf) and classification and regression trees (cart) for hotel star rating prediction. *Scientific Journal of Informatics*, *11*, 733-748. https://doi.org/10.15294/sji.v11i3.11068

Vayansky, I., & Kumar, S.A. (2020). A review of topic modeling methods. *Information Systems*, *94*, 101582. https://doi.org/10.1016/j.is.2020.101582

Wang, J., Wu, J., Sun, S., & Wang, S. (2024). The relationship between attribute performance and customer satisfaction: an interpretable machine learning approach. *Data Science and Management*, *7*(3), 164-180. https://doi.org/10.1016/j.dsm.2024.01.003

Xu, W., Yao, Z., Ma, Y., & Li, Z. (2025). Understanding customer complaints from negative online hotel reviews: a BERT-based deep learning approach. *International Journal of Hospitality Management*, *126*, 104057. https://doi.org/10.1016/j.ijhm.2024.104057

Xu, X. (2021). What are customers commenting on, and how is their satisfaction affected? examining online reviews in the on-demand food service context. *Decision Support Systems*, *142*, 113467. https://doi.org/https://doi.org/10.1016/j.dss.2020.113467

Yang, N., Korfiatis, N., Zissis, D., & Spanaki, K. (2024). Incorporating topic membership in review rating prediction from unstructured data: a gradient boosting approach. *Annals of Operations Research*, *339*, 631-662. https://doi.org/10.1007/s10479-023-05336-z

Yang, W., & Mattila, A.S. (2016). Why do we buy luxury experiences? measuring value perceptions of luxury hospitality services. *International Journal of Contemporary Hospitality Management*, *28*(9), 1848-1867. https://doi.org/10.1108/IJCHM-03-2015-0154

Yesodha, K., Anitha, R., Mala, T., & Vindhya, S. (2018). Product recommendation system using support vector machine. In *Advanced Computational and Communication Paradigms: Proceedings of International Conference on ICACCP 2017* (Vol. 1, pp. 438-446). Springer, Singapore. https://doi.org/10.1007/978-981-10-8240-5_49

Zhang, J., Du, Z., Sun, S., & Wang, S. (2025). Perception of customer satisfaction and complaints based on BERTopic and interpretable machine learning: evidence from hotels in Xi'an. *Current Issues in Tourism*, *28*(19), 3168-3190. https://doi.org/10.1080/13683500.2024.2389308

Zhang, X., Guo, F., Chen, T., Pan, L., Beliakov, G., & Wu, J. (2023). A brief survey of machine learning and deep learning techniques for e-commerce research. *Journal of Theoretical and Applied Electronic Commerce Research*, *18*(4), 2188-2216. https://doi.org/10. 3390/jtaer18040110

Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.P., Yan, H., & Li, X. (2011). Comparing twitter and traditional media using topic models. In *European Conference on İnformation Retrieval* (pp. 338-349). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-20161-5_34