Paulo Carrasco*, Pedro Esteves

# Leveraging Large Language Models for Aspect-Based Sentiment Analysis: A Restaurant Recommendation System for Entrepreneurs in Lisbon

https://doi.org/10.2478/ejthr-2026-0001
received May 8, 2025; accepted July 22, 2025

**Abstract:** The purpose of this study is to implement a restaurant recommendation system tailored for entrepreneurs within the Lisbon Metropolitan Area (AML), utilising Large Language Models (LLMs) for aspect-based sentiment analysis (ABSA) of online customer reviews. The methodology involved three phases: first, pre-processing restaurant and review data sourced from the DIG-IN online platform; second, defining relevant customer satisfaction attributes based on adapted service quality models and extracting associated keywords using LLMs; third, developing and applying prompts to GPT-4, GPT-3.5 Turbo, and Mistral 7B Instruct models to classify sentiment polarity across defined attributes like food quality, service, and ambiance found within 4,464 reviews from 115 'Italian' restaurants in the AML. Concordance tests confirmed the models' precision against human evaluation, with GPT-4o achieving 81.6% agreement. Results demonstrated the viability of using LLMs to extract, identify, and classify relevant attributes for sentiment analysis. The main conclusion is that LLMs provide a robust tool for interpreting large volumes of textual review data, enabling the creation of a prototype decision-support tool that offers entrepreneurs detailed competitive insights. This research demonstrates the practical application of LLMs and ABSA for strategic investment decisions in the highly competitive restaurant sector.

**Keywords:** Large Language Models; Online Reviews; Restaurant Industry; Aspect Based Sentiment Analysis; Natural Language Processing; Recommendation Systems

***Corresponding author: Paulo Carrasco,** Researcher at the Centre for Research, Development and Innovation in Tourism. University of Algarve. Coordinator Professor in the School of Management, Hospitality and Tourism (ESGHT-UAlg), University of Algarve. Campus da Penha. Email: pcarras@ualg.pt
**Pedro Esteves,** Master student in SME Management, School of Management, Hospitality and Tourism (ESGHT-UAlg), University of Algarve. Campus da Penha. Email: pedro.a.s.esteves@gmail.com

## 1 Introduction

Online user-generated content (UGC) is pivotal in the contemporary digital landscape, particularly within the restaurant industry. Platforms dedicated to restaurant reviews and reservations enable potential customers to access detailed information drawn from previous patrons' experiences, significantly influencing their dining choices (Asani et al., 2021). This digital word-of-mouth (WOM) constitutes a valuable resource for consumers and restaurant entrepreneurs seeking to understand customer perceptions and monitor competitors (Amaral et al., 2014). By analysing online reviews, business owners and potential investors can identify strengths and weaknesses in the competitive landscape, potentially uncovering service aspects that customers highly value but that existing establishments overlook.

However, the sheer volume of online reviews presents a significant challenge. Manually analysing thousands of comments to gauge sentiment and extract meaningful insights is often impractical and requires considerable cognitive effort and time. The evolution of computational technology, particularly in natural language processing (NLP) and generative artificial intelligence, offers robust methodologies for large-scale sentiment analysis (Jim et al., 2024). Sentiment analysis techniques aim to identify the polarity (positive, neutral, or negative) expressed within text (Ahmed et al., 2023), enabling businesses to process vast amounts of feedback efficiently.

To effectively analyse customer perceptions as reflected in online reviews, structured frameworks are beneficial. Models like DINESERV (Dining Service Quality) were developed to assess service quality across dimensions such as tangibility, reliability, responsiveness, assurance, and empathy (Stevens et al., 1995). While traditionally used with direct customer questionnaires, adapting such models is crucial for analysing the often more concise and varied nature of online feedback found on modern platforms. This adaptation is necessary to capture the specific aspects mentioned by users in digital reviews.

This study's primary objective is to develop and implement a restaurant recommendation system specifically designed for entrepreneurs and investors considering ventures in the Lisbon Metropolitan Area (AML). The system leverages Large Language Models (LLMs) to analyse customer sentiment in reviews from the DIG-IN online platform (formerly Zomato Portugal), a significant player in the Portuguese market. Specifically, the research aims to: (a) interpret and prepare restaurant and review data from DIG-IN; (b) develop an adapted sentiment analysis model based on existing literature (specifically refining DINESERV dimensions) and apply different LLMs (including versions of GPT and Mistral) for comparative analysis; and (c) present the results through a prototype tool designed to support investment decision-making in the restaurant sector.

The structure of this article proceeds logically through the research process. Section 2 provides a review of relevant literature, establishing the context regarding data-driven hospitality management, underlying technologies like Machine Learning and Large Language Models, sentiment analysis techniques including Aspect-Based Sentiment Analysis (ABSA), and established service quality frameworks such as DINESERV, alongside the rationale for adaptation. Section 3 outlines the research methodology employed, detailing the data source and pre-processing, the construction of the adapted sentiment analysis framework and attributes, the implementation and prompt engineering strategies for the LLMs, and the evaluation procedures, including concordance testing. Subsequently, Section 4 presents the key quantitative and qualitative results derived from the analyses of keyword identification, LLM performance metrics, aggregate sentiment findings for the target category, and the development of the prototype's functionalities. These results are then interpreted and contextualised in the literature and practical implications for restaurant investment within the Discussion (Section 5). Finally, Section 6 concludes the paper by summarising the primary contributions and findings, acknowledging the inherent limitations of the study, and proposing specific directions for future research in this domain.

# 2 Literature Review

## 2.1 The Ascendance of Data-Driven Management in Hospitality

Historically, business decisions, including those in the restaurant sector, often relied heavily on managerial intuition and market perception. However, technological evolution has led in an era where data-driven insights provide concrete support for strategic decision-making. Machine learning, a subset of artificial intelligence, has emerged as a powerful tool that complements human intuition by enabling analysis of vast volumes of data, surpassing the limitations of manual processing (Kozyrkov, 2018). While technology facilitates efficient data processing, human expertise remains crucial for selecting relevant data sources, interpreting results within the business context, and communicating findings effectively (Kozyrkov, 2018). Experts such as Siegel (2018) propose structured processes for implementing predictive models, emphasising collaboration between human analysts and machine capabilities, from defining objectives to operational deployment. This human-machine symbiosis is central to leveraging data science for effective management, though the increasing autonomy of AI suggests a future point where machines might play a more integrated role in governance (Schwab, 2017). Currently, however, understanding how machines process data is fundamental for utilising these paradigms effectively (Pinski et al., 2024).

## 2.2 Enabling Technologies: Machine Learning, NLP, and LLMs

Machine Learning (ML) encompasses computational techniques for extracting knowledge from large datasets, situated at the intersection of statistics, artificial intelligence, and computer science (Muller & Guido, 2016). It forms the foundation for more advanced AI applications, requiring human intervention initially to define objectives and finally to interpret results, highlighting a cooperative process (Santos, 2023). Building upon ML, Natural Language Processing (NLP) focuses on enabling computers to understand and process human language. Originating from early machine translation efforts, NLP has evolved significantly, particularly with the integration of ML techniques, enhancing its capacity for tasks like text analysis (Nadkarni et al., 2011). This capability is crucial for analysing customer sentiment expressed in reviews.

The latest advancement in this domain is the Large Language Models (LLMs), AI systems designed for direct interaction using human language. Trained on massive text datasets using deep learning techniques, LLMs like OpenAI's ChatGPT can engage in fluid conversations, interpret context, identify nuances like sarcasm, and even generate creative text (Törnberg, 2023; Huang et al., 2022). LLMs represent a significant step towards simulating human cognitive capabilities in language understanding and generation (Oliveira, 2019). Their potential to automate analytical tasks previously reliant on human interpretation is a key value proposition (Santos, 2023). The rapid adoption of generative AI across industries underscores its perceived value, with significant organisational investment expected (Chui et al., 2023). However, their use requires careful supervision to mitigate potential biases and ensure the quality of outputs (Rodrigues & Rodrigues, 2023).

## 2.3 Contextualising Restaurant Investment Research with Online Data

Recent academic literature demonstrates a growing interest in using digital data to inform restaurant location and management strategies. Studies frequently leverage data extracted from popular platforms such as Yelp.com, Koubei.com, Zomato (now DIG-IN in Portugal), and Facebook.com (Shihab & Oishi, 2018; Kim et al., 2022; Wang & Yan, 2017; Lin et al., 2016). These platforms aggregate vast amounts of user-generated reviews and restaurant information, providing rich datasets for analysis. Research spans diverse geographical locations, including major cities in the US (Wang et al., 2016), China (Wang & Yan, 2017), India (Sushmitha, 2020), Bangladesh (Ahmed et al., 2021), Singapore (Lin et al., 2016), and Portugal (Branco et al., 2024). Various analytical techniques have been employed, ranging from traditional ML algorithms like Support Vector Machines (SVM), Decision Trees, and various regression models to more advanced deep learning approaches like Deep Neural Networks (DNN) and transformer models such as BERT and RoBERTa for NLP tasks (Anmoldeep et al., 2020; Shihab & Oishi, 2018; Lin et al., 2016; Wang et al., 2016; Branco et al., 2024). Notably, the literature reviewed prior to this study indicated limited use of the latest generation of LLMs (such as GPT models) for these specific tasks, suggesting an opportunity for investigation (Törnberg, 2023).

## 2.4 Sentiment Analysis and Aspect-Based Sentiment Analysis (ABSA)

The proliferation of online review platforms has amplified the phenomenon of electronic word-of-mouth (eWOM), significantly impacting restaurant positioning and customer acquisition (Camilo, 2021; Kim et al., 2022). Sentiment analysis is a key NLP task focused on identifying the affective orientation (positive, negative, neutral) within text (Hua et al., 2023). While general sentiment analysis classifies the overall tone of a review, customer feedback is often multifaceted, expressing different sentiments towards various aspects of the experience within a single comment. For instance, a reviewer might praise the location ("lovely") but criticise the service ("a little slow"), making a single overall sentiment score insufficient for deep understanding. Aspect-Based Sentiment Analysis (ABSA) addresses this limitation by identifying sentiments associated with specific attributes or aspects mentioned in the text (Ahmed et al., 2023; Hua et al., 2024). This technique allows for a more granular analysis, decomposing reviews into specific components (e.g., food, service, ambience, price) and determining the sentiment polarity for each. ABSA provides a richer, more detailed understanding of customer perceptions, highlighting specific areas for improvement or reinforcement (Simmering & Huoviala, 2023). Given the recent advances in LLM capabilities for understanding context and nuance, they present a promising approach to performing ABSA on restaurant reviews.

## 2.5 Service Quality Models and Adaptation for Online Reviews

Measuring customer satisfaction is complex, involving subjective judgments about service experiences (Uslu & Eren, 2020). Academic research has sought to develop consistent models for this purpose. Early work by Parasuraman et al. (1988) introduced the GAP model, identifying discrepancies between customer expectations and service delivery. Subsequently, the SERVQUAL model proposed five core dimensions for service quality assessment across industries: Tangibility, Reliability, Responsiveness, Assurance, and Empathy (Parasuraman et al., 1988). Recognising the specific context of the restaurant industry, Stevens et al. (1995) adapted SERVQUAL to create DINESERV. This model retains the five core dimensions but refines the associated attributes (originally 29 items) to better reflect the dining experience. The dimensions include physical facilities and appearance (Tangibility), ability to perform the promised service dependably (Reliability), willingness to help customers and provide prompt service (Responsiveness), employee knowledge and courtesy, and their ability to inspire trust (Assurance), and caring, individualised attention (Empathy). Table 1 summarises these dimensions and illustrative criteria.

However, DINESERV was originally designed for data collection via detailed questionnaires administered immediately post-experience. Online reviews, in contrast, tend to be more concise and variable, often focusing on

**Table 1:** Dimensions and main attributes

| Dimension | Criteria |
|---|---|
| Tangibility | Physical appearance of facilities and equipment;<br>Staff uniforms and appearance;<br>Menu presentation. |
| Reliability | Preparation and service times;<br>Accuracy of orders and billing;<br>Consistency of dishes, flavour, and service. |
| Responsiveness | Staff availability;<br>Fast service;<br>Handling of special requests. |
| Assurance | Accurate and transparent information;<br>Staff training and experience;<br>Ease of interaction with staff. |
| Empathy | Sensitivity to customer needs;<br>Ability to anticipate needs;<br>Support in case of problems. |

Adapted from (Stevens et al., 1995)

overall impressions or specific salient points rather than systematically addressing all 29 DINESERV items. For example, multiple DINESERV items relating to physical aspects (e.g., attractive décor, comfortable seating, cleanliness) might be condensed into a single general comment about the "ambience" or "cleanliness" in an online review. Therefore, for effective analysis using LLMs on online review data, adapting the DINESERV framework is necessary. This involves consolidating the original items into broader, more representative criteria that capture the essence of each dimension while aligning with the language typically used in online comments. This study undertook such an adaptation, drawing on the DINESERV dimensions and related literature on key restaurant choice attributes (e.g., Choi et al., 2009; Haghighi, 2012; Alves, 2017; Silva et al., 2009; Alhelalat et al., 2017). The goal was to create a synthesized set of attributes linked to the core DINESERV dimensions, suitable for guiding the LLM-based ABSA process. This adapted framework (to be presented in section 3.4) provides a structured yet flexible approach to analysing sentiment expressed in concise online reviews, forming the basis for the methodology employed in this research.

# 3 Methodology

This study implements a restaurant recommendation system for the Lisbon Metropolitan Area (AML) using sentiment analysis of online reviews, designed as a decision-support tool for potential investors. The methodology detailed below outlines the data acquisition, pre-processing, sentiment analysis framework development, Large Language Model (LLM) implementation strategies, and the model evaluation procedures.

## 3.1 Data Source and Acquisition

The primary data for this research were provided by ZMT-EUROPE, LDA, the company operating the DIG-IN digital restaurant discovery platform (previously Zomato Portugal). Data were obtained through the company's Datalab initiative. The dataset encompasses two main components pertinent to this study: (1) information on restaurants located within the AML, including characteristics like cuisine type, location, and pricing, and (2) user-generated content associated with these restaurants, specifically numerical ratings and textual reviews (comments). This collaboration provided access to a substantial and relevant dataset reflecting real-world customer experiences within the target geographical area.

## 3.2 Data Characteristics and Pre-processing

The initial dataset contained information on 15,892 restaurants within the AML and 1,043,806 associated user ratings, of which 474,470 included written comments. Given the large volume and potential for inconsistencies (e.g., null values, inconsistent geographical definitions, and varied cuisine classifications), data pre-processing was performed using the Python programming language and associated data analysis libraries. This phase involved cleaning the data, ensuring correct geographical delimitation of the AML, and standardising restaurant information. Two primary datasets resulted from this process:

– A restaurant dataset containing unique identifiers (restaurant_id), name, cuisine type(s), average rating, price level, location coordinates, operational features (e.g., outdoor seating, takeaway), and review count.
– A reviews dataset linking user comments (original_text) to specific restaurants (restaurant_id) and including the overall rating assigned by the user (general_rating).

## 3.3 Sample Selection for Prototype Development

Due to the potential diversity and inconsistency in how restaurants self-classify their cuisine types on the platform, an initial analysis identified the ten most frequent cuisine categories within the AML dataset. To ensure a focused yet representative analysis for the prototype development and manage computational costs associated with LLM processing, a specific cuisine category was selected. The 'Italian' category was chosen based on three key factors derived from analyzing restaurants offering only one cuisine type: (1) a favorable comment-to-restaurant ratio (average 38.82 comments per restaurant), indicating sufficient opinion density; (2) a substantial number of establishments (115 restaurants), ensuring reasonable representativeness; and (3) a manageable total number of comments (4,464 written comments) that allowed for complete processing within practical resource constraints, unlike categories with significantly higher comment volumes (e.g., 'Burger'). This selection provided an ideal balance between data richness and processing feasibility for demonstrating the prototype's capabilities.

## 3.4 Defining a Framework for Sentiment Analysis

To guide the sentiment analysis, a framework was developed by adapting established service quality models to the specific context of online reviews. Building upon the five dimensions of the DINESERV model (Tangibility, Reliability, Responsiveness, Assurance, Empathy) (Stevens et al., 1995), which are often too detailed for concise online comments, the twenty-nine original items were consolidated into a set of ten attributes frequently discussed in restaurant-selection literature and relevant to customer experience (Choi et al., 2009). An exploratory assessment of the present dataset showed that the average comment contains approximately sixty words, confirming that reviewers rarely address all twenty-nine items systematically and thereby reinforcing the need for aggregation.

The resulting attributes are Food Taste, Food Portion, Service, Price, Ambience / Atmosphere, Food Presentation, Nutritious Food, Restaurant Reputation, Cleanliness and Variety of Healthy Meals (Table 2). This structure preserves the theoretical alignment with the five core dimensions while adopting labels that mirror the vocabulary commonly found in user-generated content.

Then, an LLM (ChatGPT 3.5) was utilised via a specific prompt to assist in mapping these ten attributes to the expanded ten attributes, ensuring that the analysis framework captured the essence of service quality while

**Table 2:** Consolidated attributes for Aspect-Based Sentiment Analysis

| Attribute | Description |
|---|---|
| 1. Food Taste | Assesses the quality of food flavour, including authenticity and balance of ingredients. |
| 2. Food Portion | Refers to the quantity served, considering appropriateness in terms of price and expectations. |
| 3. Service | Seeks speed, efficiency, courtesy, and professionalism of the staff. |
| 4. Price | Evaluates the adequacy of prices in relation to the quality of dishes and service. |
| 5. Ambiance/Atmosphere | Considers decoration, lighting, music, and spatial arrangement. |
| 6. Food Presentation | Observes visual preparation, care, and creativity in the presentation of dishes. |
| 7. Nutritious Food | Analyses the nutritional quality of food options, such as the use of fresh ingredients and preparation methods. |
| 8. Restaurant Reputation | Based on reviews, recommendations, awards received, and public recognition of the restaurant. |
| 9. Cleanliness | Assess the hygiene of the restaurant, including dining and food preparation areas. |
| 10. Variety of Healthy Meals | Measures the diversity of healthy options available on the menu, as well as other menu alternatives. |

Adapted from (Choi et al., 2009)

focusing on aspects commonly mentioned by online reviewers. This adapted framework provides the structure for the aspect-based sentiment analysis (ABSA).

## 3.5 LLM Implementation and Prompt Engineering

The core sentiment analysis was performed using LLMs accessed via the openrouter.ai platform (https://openrouter.ai), which provides a unified API for interacting with various models.

### 3.5.1 Keyword Identification

To aid attribute identification within online reviews, an initial methodological step involved using an LLM (ChatGPT 3.5) to analyse a sample of 1,000 comments. This was guided by Prompt A (see Appendix 1), designed specifically to elicit relevant keywords associated with each of the ten defined attributes. The list of keywords and their frequencies generated by this process served as input to the subsequent attribute classification process. The specific keywords identified and their resulting frequencies are presented in the Results section (Section 4.1).

### 3.5.2 LLM Selection and Prompt Design

Three LLMs were selected for processing and evaluation based on their capabilities, accessibility, and cost-effectiveness:

– ChatGPT 3.5 Turbo (openai/gpt-3.5-turbo): A widely used paid model known for its balance of efficiency and cost.
– ChatGPT 4o_2024-05-13 (openai/gpt-4o-2024-05-13): A more advanced paid model recognised for higher precision, particularly in complex tasks.
– Mistral 7B Instruct (mistralai/mistral-7b-instruct: free): A freely available model demonstrating competitive performance in testing, offering a cost-effective alternative.

A detailed prompt was engineered to instruct the LLMs on performing the ABSA task consistently. The prompt (presented in Appendix 2 – Prompt B) specified the objective (classify experience based on written comments), defined the ten attributes of evaluation with associated keywords and descriptions, and outlined the instructions for analysis. Key instructions included: assigning a 1-5 Likert scale score (1=very dissatisfied, 5=very satisfied) to each attribute identified in the comment, using adjectives and keywords to infer satisfaction levels, assigning ‘N/A’ if an attribute was not mentioned, and providing the output in a specific structured format listing the classification for each attribute. This structured prompt aimed to ensure consistent and comparable outputs across different models and comments.

## 3.6 Model Evaluation

Several steps were taken to evaluate the accuracy and consistency of the selected LLMs in performing the sentiment analysis task according to the defined framework.

### 3.6.1 Initial Human vs. Machine Comparison

A preliminary test compared the overall sentiment classification (1-5 stars) assigned by ChatGPT 3.5 to that assigned by the original human reviewer for 50 comments sampled from TripAdvisor. This initial check showed high agreement (47 out of 50 matched), providing early confidence in the model’s ability to interpret overall sentiment from text.

### 3.6.2 Inter-Model Concordance

The consistency among the three selected LLMs (GPT-3.5 Turbo, GPT-4o, Mistral 7B Instruct) in their attribute classification was assessed. A sample of 100 comments was processed by all three models using Prompt B. Concordance was measured based on whether the models agreed on the sentiment polarity (positive: ratings 3-5; negative: ratings 1-2) or agreed on the absence of an attribute for each of the ten attributes. This procedure aimed to understand the reliability across different models. The specific concordance rates are detailed in Section 4.2.2.

### 3.6.3 Human vs. Machine Concordance (Extended)

A more rigorous concordance analysis was performed by comparing the attribute classifications produced by each of the three LLMs with those of a single human expert evaluator (serving as the reference) for a random sample of 100 comments from the ‘Italian’ dataset. Concordance was defined as agreement on the presence and polarity (positive/negative) of an attribute, or agreement on its absence. The detailed performance results for each model against the human expert are presented in Table 4 in Section 4.2.3.

# 4 Results

This section presents the findings from applying the methodology, including keyword analysis, evaluation of Large Language Model (LLM) performance in sentiment classification, aggregate analysis of the selected restaurant category, and the functionality of the developed prototype tool.

## 4.1 Attribute Keyword Analysis

The methodological step of identifying keywords associated with the ten defined customer experience attributes (detailed in Section 3.4, Table 2), using LLM analysis (Prompt A, Appendix 1) on a sample of 1,000 comments, revealed distinct frequency patterns. These are summarised in Table 3.

Keywords related to Service (312 mentions), Price (264 mentions), Ambiance/atmosphere (217 mentions), Taste of the food (112 mentions), and Food Portion (119 mentions) were most prevalent. This indicates that customers frequently focus their feedback on these core aspects of the dining experience. Conversely, attributes such as Reputation of the restaurant (19 mentions), Cleanliness (7 mentions), Presentation of the food (45 mentions), and Variety of healthy meals (53 mentions) were mentioned less explicitly. Despite the low frequency for some, all ten attributes were retained for subsequent analysis to ensure comprehensive coverage.

## 4.2 LLM Concordance and Performance Evaluation

The performance of the selected LLMs (ChatGPT 3.5 Turbo, ChatGPT 4o_2024-05-13, Mistral 7B Instruct) was evaluated through multiple concordance tests as outlined in the methodology (Section 3.6).

### 4.2.1 Initial Sentiment Agreement

The initial comparison using 50 TripAdvisor comments (Section 3.6.1) found that ChatGPT 3.5 assigned the same overall numerical rating (1-5 stars) as the human reviewer in 47 out of 50 cases (94% agreement), demonstrating a strong capability to infer general sentiment from text alone.

### 4.2.2 Inter-Model Attribute Concordance

When analysing the classification polarity or absence of attributes across 100 comments (Section 3.6.2), the three LLMs showed concordance rates of 80% or higher for five key attributes: Taste of the food, Service, Price, Ambi-

**Table 3:** List of keywords and frequencies

| Attribute | Keywords | Frequency |
|---|---|---|
| 1. Food Taste | tasty, delicious, seasoning, authentic, nice, balanced… | 112 |
| 2. Food Portion | portion, quantity, sufficient, generous, large… | 119 |
| 3. Service | service, quick, efficient, friendly, kind, professional… | 312 |
| 4. Price | price, cost, expensive, cheap, affordable… | 264 |
| 5. Ambiance/Atmosphere | ambience, atmosphere, cosy, decoration, lighting, music… | 217 |
| 6. Food Presentation | presentation, visual, plating, creative, decorated… | 45 |
| 7. Nutritious Food | nutritious, healthy, fresh, natural ingredients… | 98 |
| 8. Restaurant Reputation | famous, recommended, awarded, well-known, reputation… | 19 |
| 9. Cleanliness | clean, hygienic, tidy, cleanliness… | 7 |
| 10. Variety of Healthy Meals | variety, healthy options, healthy menu, alternatives… | 53 |

**Table 4:** Agreement Rates Between Models and Human Evaluator

| LLM model | Agreement Rate |
|---|---|
| ChatGPT 3.5 Turbo | 64,4% |
| ChatGPT 4o_2024-05-13 | 81,6% |
| Mistral 7B Instruct Free | 60,6% |

ance/atmosphere, and Variety of healthy meals. This high agreement aligns with the high frequency of keywords identified for these attributes. The lowest inter-model concordance was observed for Food Portion (55.0%), potentially indicating greater ambiguity or less standardised language used by reviewers for this attribute.

#### 4.2.3 Human vs. Machine Attribute Concordance

The comparison against expert human evaluation on 100 comments (Section 3.6.3) revealed the performance differences presented in Table 4.

ChatGPT 4o_2024-05-13 achieved the highest concordance rate with the human evaluator at 81.6%. This was substantially higher than ChatGPT 3.5 Turbo (64.4%) and Mistral 7B Instruct (60.6%). The 81.6% concordance rate for GPT-4o approaches typical inter-human agreement levels reported in related studies (Carrasco & Dias, 2024), indicating its strong ability to replicate human-like judgment in identifying and classifying sentiment towards specific attributes. Analysis of specific discrepancies highlighted the inherent subjectivity in sentiment analysis, even between humans, but confirmed the general robustness of the LLMs, particularly GPT-4o.

### 4.3 Sentiment Analysis of 'Italian' Restaurant Reviews

The entire set of 4,464 comments for the 115 'Italian' restaurants in the AML (selected as per Section 3.3) was processed using the ChatGPT 3.5 Turbo model (guided by Prompt B, Appendix 2) to generate attribute-level sentiment scores. A comparison between the overall sentiment rating derived from the model's attribute analysis and the customer's original rating showed a high degree of alignment. Specifically, 83.96% of the comments processed had a difference of 1.0 Likert scale points or less between the customer's rating and the model's derived rating. This finding suggests that even the less advanced GPT-3.5 Turbo model demonstrated considerable competence in capturing the overall sentiment expressed across the various attributes mentioned in the reviews.

### 4.4 Prototype Tool Functionality and Insights

The sentiment analysis results derived from processing the 4,464 comments were integrated into an interactive prototype tool. This tool is designed to provide potential investors with actionable insights through two main visual components:

The first component is an Interactive Map, which visualises a selected geographical area (e.g., Marquês de Pombal, as shown in Figure 1a). This map displays the locations of competing restaurants within the chosen category ('Italian') and employs features like a colour grid to indicate market saturation (density of competitors). Users can click on individual restaurant pins to access detailed pop-ups containing basic information and summarised sentiment analysis results (Figure 1b). This interface provides a clear and immediate overview of the competitive landscape in the selected vicinity.

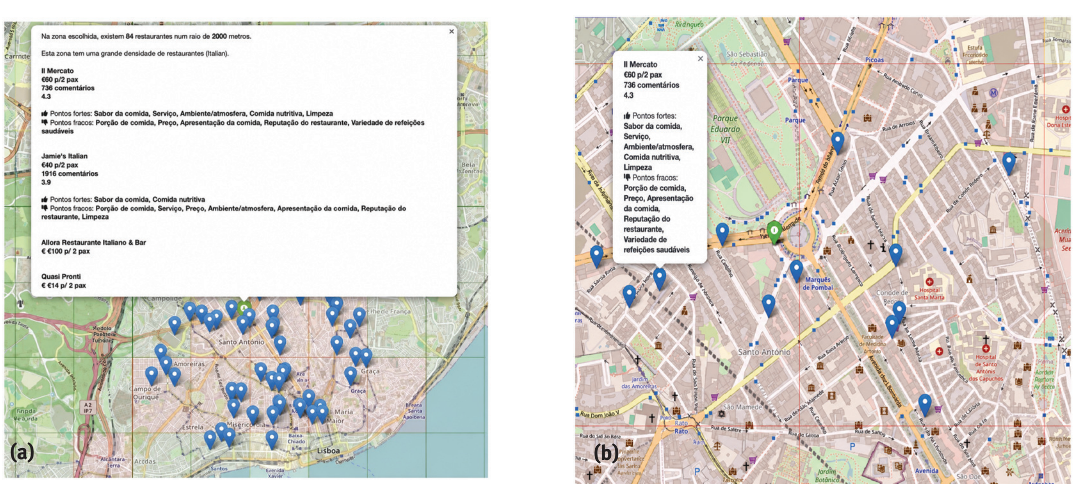


**Figure 1:** Prototype visual components (a) map of the locations of competing restaurants within a chosen category; (b) information and summarised sentiment analysis results for a specific restaurant

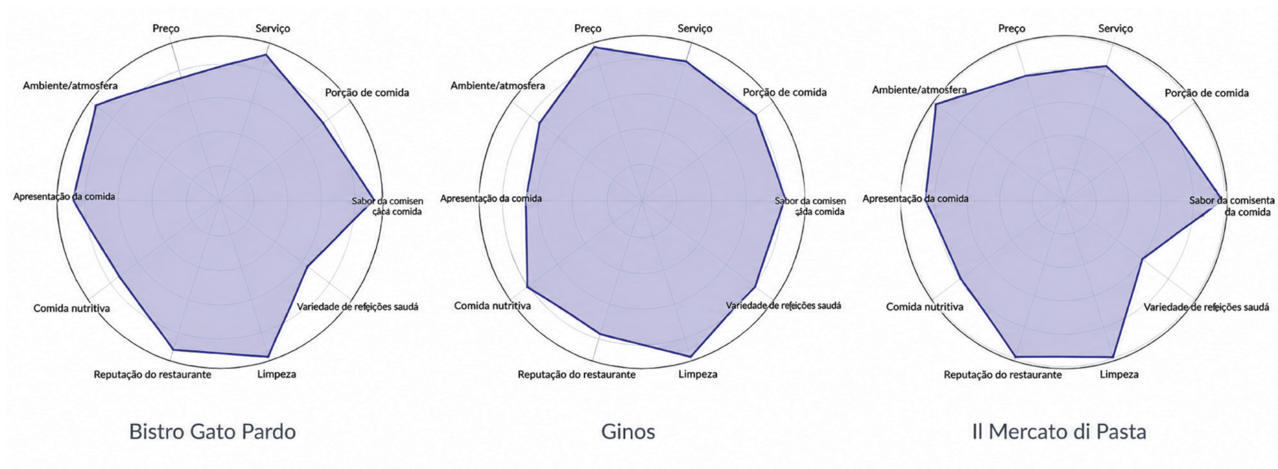


**Figure 2:** Radar chart comparison of multiple restaurants

The second component features Restaurant Performance Radar Charts. For selected competitors, the tool generates radar charts displaying the average sentiment score (derived from LLM analysis of reviews) for each of the ten defined attributes (e.g., Taste, Service, Price, Ambience), as illustrated in Figure 2. This allows investors to quickly visualise and compare the perceived strengths and weaknesses of competitors based on customer feedback, moving beyond simple overall ratings to understand performance on specific aspects crucial to customer experience.

Together, these features centralise competitive information and sentiment analysis, enabling investors to identify market opportunities and potential areas for differentiation based on detailed, data-driven insights derived directly from customer comments.

## 5 Discussion

This study's exploration into leveraging Large Language Models (LLMs) for aspect-based sentiment analysis (ABSA) of online restaurant reviews, aimed at informing investment decisions, generated favourable insights into the practical application and capabilities of these technologies within the hospitality sector. The findings highlight that contemporary LLMs, particularly GPT-4o, can interpret and classify sentiment towards specific dining attributes with a high degree of accuracy, as evidenced by the 81.6% concordance rate with expert human evaluation (Table 4). This performance, which approaches inter-human agreement levels (Carrasco & Dias, 2024), signifies a critical step towards reliably automating the nuanced task of ABSA. Such automation is important for overcoming the scalability challenges associated with manual review analysis (Jim et al., 2024) and represents an evolution in LLM utility from text generation to sophisticated analytical interpretation (Törnberg, 2023).

The successful application of ABSA was significantly supported by the adapted analytical framework, which synthesised established DINESERV dimensions (Stevens et al., 1995) with ten specific attributes pertinent to restaurant selection (Choi et al., 2009), detailed in Table 2. The keyword frequency analysis (Table 3) confirmed that this framework captured aspects commonly discussed by customers, particularly concerning service, price, ambience, and food quality. By enabling a granular dissection of customer perceptions (Ahmed et al., 2023; Hua et al., 2024), this approach provides deeper insights than general sentiment scores, which, while showing broad agreement in the large-scale GPT-3.5 analysis (Results 4.3), can mask critical details about specific strengths or weaknesses.

The practical implications of integrating LLM-driven sentiment analysis with market data are demonstrated by the developed prototype tool (Figures 1 and 2). This tool operationalises data-driven decision-making principles (Kozyrkov, 2018) by offering investors a clear and intuitive means to assess competitor performance across specific attributes and identify market opportunities (Bressler, 2012). For instance, by using the radar charts (Figure 2) to detect consistent weaknesses in competitor offerings, such as poor service reviews or negative sentiment on food portions despite good taste, investors can strategically tailor their own concepts to fill these identified gaps. This capability to move beyond basic competitive intelligence (location, price) to understand nuanced customer perceptions offers a distinct advantage in planning and launching new ventures in a saturated market.

Despite the robust performance, particularly of GPT-4o, the study also acknowledged the inherent subjectivity in sentiment analysis. Discrepancies between LLM and human classifications, especially for reviews with mixed emotions or sarcasm (Results 4.2.2), and the 16.04% of comments showing a sentiment difference greater than one Likert point in the large-scale analysis (Results 4.3), underscore this challenge. However, given that perfect agreement is rare even among human evaluators, the consistency achieved by the LLMs in navigating this subjectivity and extracting meaningful trends from diverse, real-world commentary is noteworthy. This research, therefore, validates the use of modern LLMs, guided by a contextually adapted analytical framework, to conduct effective ABSA on online restaurant reviews, paving the way for more sophisticated, AI-driven competitive analysis in the hospitality industry.

## 6 Conclusions

This research fulfilled its primary objective of developing a system that utilises Large Language Models (LLMs) for aspect-based sentiment analysis (ABSA) of online restaurant reviews. The outcome was a prototype tool designed to support investment decisions within the Lisbon Metropolitan Area (AML) by facilitating detailed competitor analysis. A key conclusion from this study is that LLMs can effectively automate the extraction and classification of sentiment towards specific customer experience attributes from extensive textual data. The models demonstrated a significant ability to interpret customer feedback when guided by engineered prompts based on adapted service quality literature.

The significance of this research lies in its practical application to Business Intelligence and applied Natural Language Processing within the hospitality sector. It offers a data-driven methodology for navigating the competitive restaurant industry. The developed prototype serves as proof-of-concept, illustrating how such a system can aid in enhancing investment decisions, identifying market opportunities, and potentially contributing to the sustainability of new ventures by transforming unstructured reviews into structured insights.

Several limitations are acknowledged. The prototype's current scope is restricted to a single data source and cuisine category within the AML. The attribute framework, though informed by literature, may benefit from further adaptation to evolving online customer discourse. The initial keyword identification was based on a limited sample, and interpreting complex linguistic nuances remains an ongoing challenge for LLMs. Additionally, the computational costs of using certain LLMs warrant consideration for widespread application.

These limitations inform suggestions for future research. Expanding data sources and geographical coverage would improve the system's applicability. Continued investigation into pertinent customer attributes and keyword refinement is recommended. Further advancements in LLM training for nuanced language and prompt optimisation, including the exploration of open-source models, could enhance the system's robustness, efficiency, and accessibility for strategic decision-making in the hospitality industry.

## Disclosure statement

This is to acknowledge any financial interest or benefit that has arisen from the direct applications of your research.

## Conflicts of Interest

The authors of the article "Leveraging Large Language Models for Aspect-Based Sentiment Analysis: A Restaurant Recommendation System for Entrepreneurs in Lisbon" declare no conflict of interest.

## Bionotes

**Paulo Carrasco** is a Coordinating Professor at ESGHT-UAlg and a researcher at CiTUR Algarve. He holds a PhD in Applied Mathematics to Economics and Management from ISEG – University of Lisbon. His research interests include the application of artificial intelligence and mathematics to management and education. ORCID: 2A14-9818-5274

**Pedro Esteves** holds a master's degree in Small and Medium-Sized Enterprise (SME) Management. His research interests focus on the application of artificial intelligence to entrepreneurship and management.

## References

[1] Ahmed, K., Nadeem, M., Zheng, Z., Li, D., Ullah, I., Assam, M., Ghadi, Y., & Mohamed, H. (2023). Breaking down linguistic complexities: A structured approach to aspect-based sentiment analysis. *Journal of King Saud University – Computer and Information Sciences, 35*, Article. 101651. https://doi.org/10.1016/j.jksuci.2023.101651

[2] Ahmed, T., Akhter, L., Talukder, F., Al-Monsur, H., Rahman, H., & Sattar, A. (2021). Restaurant recommendation system in Dhaka City using machine learning approach. *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)* (pp. 59–63). IEEE. https://doi.org/10.1109/SMART52563.2021.9676197

[3] Alhelalat, J., Habiballah, M., & Twaissi, N. (2017). The impact of personal and functional aspects of restaurant employee service behaviour on customer satisfaction. *International Journal of Hospitality Management, 66*, 46–53. https://doi.org/10.1016/j.ijhm.2017.07.001

[4] Alves, C. (2017). Estou Satisfeito: Um Estudo sobre a Lealdade do Consumidor em Restaurantes Empregando a Escala Dineserv. *Revista Brasileira de Marketing, 16*(3), 334–350. https://doi.org/10.5585/remark.v16i3.3389

[5] Amaral, F., Tiago, T., & Tiago, F. (2014). User-generated content: Tourists' profiles on TripAdvisor. *International Journal of Strategic Innovative Marketing, 1*(3), 165-170. https://doi.org/10.15556/IJSIM.01.03.002

[6] Anmoldeep, A., Spoorthy, D., Jayanth, G., Jnanashree, A., & Sushmitha, S. (2020). Geographical location recommendation system for restaurants in Bengaluru. *Journal of Emerging Technologies and Innovative Research, 7*(6), 2111–2114. https://www.jetir.org/papers/JETIR2006306.pdf

[7] Asani, E., Vahdat-Nejad, H., & Sadri, J. (2021). Restaurant recommender system based on sentiment analysis. *Machine Learning with Applications, 6*, Article 100114. https://doi.org/10.1016/j.mlwa.2021.100114

[8] Branco, A., Parada, D., Silva, M., Mendonça, F., Mostafa, S., & Dias, F. (2024). Sentiment analysis in Portuguese restaurant reviews: Application of transformer models in edge computing. *Electronics, 13*(3), Article 589. https://doi.org/10.3390/electronics13030589

[9] Bressler, M. (2012). How small businesses master the art of competition through superior competitive advantage. *Journal of Management & Marketing Research, 11*, 30–41.

[10] Camilo, J. (2021). *Análise de sentimentos: Relação da experiência do consumidor com base nos problemas evidenciados nas online reviews* [Master's dissertation, Universidade do Porto]. Repositório Aberto da Universidade do Porto. https://hdl.handle.net/10216/134916

[11] Carrasco, P., & Dias, S. (2024). Enhancing restaurant management through aspect-based sentiment analysis and NLP techniques. *Procedia Computer Science, 237*, 129–137. https://doi.org/10.1016/j.procs.2024.05.088

[12] Choi, J., Lee, J., & Zhao, J. (2009, July 29-August 1). *A Comparison of the Restaurant Selection Preferences between Residents and Visitors of South*. International CHRIE Conference-Refereed Track, San Francisco, CA, United States. https://scholarworks.umass.edu/refereed/Sessions/Friday/14/

[13] Chui, M., Yee, L., Hall, B., Singla, A., & Sukharevsky, A. (2023, August 1). *The state of AI in 2023: Generative AI's breakout year*. McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year

[14] Pinski, M., Hofmann, T., & Benlian, A. (2024). AI literacy for the top management: An upper echelons perspective on corporate AI orientation and implementation ability. *Electronic Markets*, 34(1), 1–24. https://doi.org/10.1007/s12525-024-00707-1

[15] Colson, E. (2019, July-August). What AI-driven decision making looks like. *Harvard Business Review*. https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like

[16] Haghighi, M. (2012). Evaluation of factors affecting customer loyalty in the restaurant industry. *African Journal of Business Management, 6*(14), 5039-5046. https://doi.org/10.5897/AJBM11.2765

[17] Hua, Y., Denny, P., Taskova, K., & Wicker, J. (2023). *A Systematic Review of Aspect-based Sentiment Analysis (ABSA): Domains, Methods, and Trends*. arXiv. https://doi.org/10.48550/arXiv.2311.10777

[18] Huang, J., Gu, S., Hou, L., Wu, Y., Wang, X., Yu, H., & Han, J. (2022). *Large Language Models Can Self-Improve*. arXiv. https://doi.org/10.48550/arXiv.2210.11610

[19] Jim, J., Talukder, M., Malakar, P., Kabir, M., Nur, K., & Mridha, M. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal, 6*, Article 100059. https://doi.org/10.1016/j.nlp.2024.100059

[20] Kim, J., Lee, M., Kwon, W., Park, H., & Back, K. (2022). Why am I satisfied? See my reviews – Price and location matter in the restaurant industry. *International Journal of Hospitality Management, 101*, Article 103111. https://doi.org/10.1016/j.ijhm.2021.103111

[21] Kozyrkov, C. (2018, December 4). *What Great Data Analysts Do—And Why Every Organization Needs Them*. Harvard Business Review. https://hbr.

org/2018/12/what-great-data-analysts-do-and-why-every-organization-needs-them

[22] Lin, J., Oentaryo, R., Lim, E., Vu, C., Vu, A., & Kwee, A. (2016). Where is the goldmine? Finding promising business locations through Facebook data analytics. *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 93–102). IEEE. https://doi.org/10.1109/ASONAM.2016.7752178

[23] Muller, A., & Guido, S. (2016). *Introduction to Machine Learning with Python* (1st ed.). O'Reilly Media.

[24] Nadkarni, P., Ohno-Machado, L., & Chapman, W. (2011). Natural language processing: An introduction. *Journal of the American Medical Informatics Association, 18*(5), 544–551. https://doi.org/10.1136/amiajnl-2011-000464

[25] Oliveira, A. (2019). *Inteligência artificial* (1ª ed.). Fundação Francisco Manuel dos Santos.

[26] Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64*(1), 12–40.

[27] Rodrigues, O., & Rodrigues, K. (2023). A inteligência artificial na educação: Os desafios do ChatGPT. *Texto Livre: Linguagem e Tecnologia, 16*, Article e45997. https://doi.org/10.1590/1983-3652.2023.45997

[28] Santos, A. (2023a). Desafios e Oportunidades da Inteligência Artificial na Educação e na Formação. *RE@D - Revista de Educação a Distância e Elearning, 6*(2), Article 2. https://doi.org/10.34627/redvol6iss2e202311

[29] Santos, A. (2023b). O que é, em que consiste o Chat GPT? *RE@D - Revista de Educação a Distância e Elearning, 6*(2). https://doi.org/10.34627/redvol6iss2e202311

[30] Schwab, K. (2017). *A Quarta Revolução Industrial*. LEVOIR, Marketing e Conteúdos Multimédia, S.A.

[31] Shihab, I., & Oishi, M. (2018). A machine learning approach to suggest ideal geographical location for new restaurant establishment. *2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)* (pp. 1-6). IEEE. https://doi.org/10.1109/R10-HTC.2018.8629845

[32] Siegel, E. (2018, October 3). *3 Common Mistakes That Can Derail Your Team's Predictive Analytics Efforts*. Harvard Business Review. https://hbr.org/2018/10/3-common-mistakes-that-can-derail-your-teams-predictive-analytics-efforts

[33] Silva, L., Medeiros, C., & Costa, B. (2009). Qualidade dos Serviços Turísticos no Setor de Restaurantes: Uma Aplicação do Modelo SERVPERF. *Revista Hospitalidade, 6*(Edição Especial), 115–139.

[34] Simmering, P., & Huoviala, P. (2023). *Large language models for aspect-based sentiment analysis*. arXiv. https://arxiv.org/abs/2310.18025

[35] Statista. (2022). *Python use cases worldwide 2022*. https://www.statista.com/statistics/1338409/python-use-cases/

[36] Statista. (2024a). *Languages most frequently used for web content as of January 2024, by share of websites*. https://www.statista.com/statistics/262946/most-common-languages-on-the-internet/

[37] Statista. (2024b). *Most used programming languages among software developers worldwide as of 2024*. https://www.statista.com/statistics/793628/worldwide-developer-survey-most-used-languages/

[38] Statista. (2024c). *Number of monthly active Facebook users worldwide as of 4th quarter 2023*. https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

[39] Stevens, P., Knutson, B., & Patton, M. (1995). Dineserv: A tool for measuring service quality in restaurants. *The Cornell Hotel and Restaurant Administration Quarterly, 36*(2), 56–60. https://doi.org/10.1016/0010-8804(95)93844-K

[40] Sushmitha, S. (2020). Geographical location recommendation system for restaurants in Bengaluru. *Journal of Emerging Technologies and Innovative Research (JETIR), 7*(6), 2111–2114. https://www.jetir.org/papers/JETIR2006306.pdf

[41] Törnberg, P. (2023). *How to use LLMs for Text Analysis*. arXiv. https://arxiv.org/abs/2307.13106

[42] Uslu, A., & Eren, R. (2020). Critical review of service quality scales with a focus on customer satisfaction and loyalty in restaurants. *DETUROPE - The Central European Journal of Tourism and Regional Development, 12*(1), 64–84. https://doi.org/10.32725/det.2020.004

[43] Wang, F., Chen, L., & Pan, W. (2016). Where to place your next restaurant? Optimal restaurant placement via leveraging user-generated reviews. *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (pp. 2371–2376). ACM. https://doi.org/10.1145/2983323.2983696

[44] Wang, J., & Yan, X. (2017). How location matters in restaurant success? *Proceedings of the Pacific Asia Conference on Information Systems (PACIS) 2017*, Paper 237. https://aisel.aisnet.org/pacis2017/237/

## Appendix 1 - Prompt A

*"You are a sentiment analyst who reads restaurant reviews and classifies attributes based on customer experience. In this task, you must focus on the 10 attributes listed below and identify, in the 1000 reviews contained in the attached file, the ideal keywords to best capture each of the attributes.*

1. *Food Taste – Assesses the quality of food flavour, including authenticity and balance of ingredients.*
2. *Food Portion – Refers to the quantity served, considering appropriateness in terms of price and expectations.*
3. *Service – Seeks speed, efficiency, courtesy, and professionalism of the staff.*
4. *Price – Evaluates the adequacy of prices in relation to the quality of dishes and service.*
5. *Ambiance/Atmosphere – Considers decoration, lighting, music, and spatial arrangement.*
6. *Food Presentation – Observes visual preparation, care, and creativity in the presentation of dishes.*
7. *Nutritious Food – Analyses the nutritional quality of food options, such as the use of fresh ingredients and preparation methods.*
8. *Restaurant Reputation – Based on reviews, recommendations, awards received, and public recognition of the restaurant.*
9. *Cleanliness – Assesses the hygiene of the restaurant, including dining and food preparation areas.*
10. *Variety of Healthy Meals – Measures the diversity of healthy options available on the menu, as well as other menu alternatives.*

*After listing the keywords for each attribute, I want you to count the incidence of each attribute in the reviews.*
*Organize a table with one column for the attributes, one column for the identified keywords, and one column for the number of times you detect the attributes in the reviews."*

## Appendix 2 - Prompt B

*"You are a restaurant review evaluator.*
*Objective: To classify restaurant experiences based on comments written by users.*

*Evaluation Attributes:*
1. *Food Taste: Assesses the quality of food flavour, including authenticity and balance of ingredients.*
2. *Food Portion: Refers to the quantity served, considering appropriateness in terms of price and expectations.*
3. *Service: Looks for speed, efficiency, friendliness, and professionalism of the staff.*
4. *Price: Evaluates the adequacy of prices relative to the quality of dishes and service.*
5. *Ambiance/Atmosphere: Considers decoration, lighting, music, and the spatial arrangement of the venue.*
6. *Food Presentation: Observes the visual preparation, care, and creativity in the presentation of dishes.*
7. *Nutritious Food: Analyses the nutritional quality of food options, including the use of fresh ingredients and preparation methods.*
8. *Restaurant Reputation: Based on reviews, recommendations, awards received, and public recognition of the restaurant.*
9. *Cleanliness: Assesses the hygiene of the restaurant, including both dining and food preparation areas.*
10. *Variety of Healthy Meals: Measures the diversity of healthy options available on the menu, as well as other menu alternatives.*

*Example Keywords per Attribute:*
1. *Food Taste: tasty, delicious, seasoning, authentic, nice, balanced*
2. *Food Portion: portion, quantity, sufficient, generous, large*
3. *Service: service, fast, efficient, friendly, kind, professional*
4. *Price: price, cost, expensive, cheap, affordable*
5. *Ambiance/Atmosphere: ambience, atmosphere, cosy, decoration, lighting, music*
6. *Food Presentation: presentation, visual, plating, creative, decorated*
7. *Nutritious Food: nutritious, healthy, fresh, natural ingredients*

8. *Restaurant Reputation: famous, recommended, awarded, well-known, reputation*
9. *Cleanliness: clean, hygienic, tidy, cleanliness*
10. *Variety of Healthy Meals: variety, healthy options, healthy menu, alternatives*

*Instructions:*
*Use the adjectives mentioned in the review to help infer the satisfaction level for each attribute.*

– *Assign a rating from 1 to 5 for each of the listed criteria, based on the descriptions provided in the review;*
– *If you cannot identify any words that allow classification of an attribute, mark it as N/A.*
– *Expected Output Format:*
  *Food Taste: <rating>*
  *Food Portion: <rating>*
  *Service: <rating>*
  *Price: <rating>*
  *Ambiance/Atmosphere: <rating>*
  *Food Presentation: <rating>*
  *Nutritious Food: <rating>*
  *Restaurant Reputation: <rating>*
  *Cleanliness: <rating>*
  *Variety of Healthy Meals: <rating>"*