Sentiment Analysis for Hotel Reviews: A Systematic Literature Review

3 authors:

Asma Ameur
Tunisia Polytechnic School
5 PUBLICATIONS   **52** CITATIONS

SEE PROFILE

Sana Hamdi
Tunis El Manar University
16 PUBLICATIONS   **171** CITATIONS

SEE PROFILE

Sadok Ben Ben Yahia
University of Southern Denmark
523 PUBLICATIONS   **4,776** CITATIONS

SEE PROFILE

Sentiment Analysis for Hotel Reviews: A Systematic Literature Review

ASMA AMEUR, Polytechnic School of Tunisia & LIPAH, LR111417, Tunis El Manar University, Faculty of Sciences of Tunis, Tunisia

SANA HAMDI, LIPAH, LR111417, Tunis El Manar University, Faculty of Sciences of Tunis, Tunisia

SADOK BEN YAHIA, Department of Software Science, Tallinn University of Technology, Estonia

Sentiment Analysis (SA) helps to automatically and meaningfully discover hotel customers' satisfaction from their shared experiences and feelings on social media. Several studies have been conducted to improve the precision of SA in the hospitality industry, which vary in data preprocessing techniques, feature representation, sentiment classification levels, and models, and they use different datasets. Such variations are worthy of attention and monitoring. Despite the importance of SA in hospitality and tourism, review studies identifying gaps and suggesting future research directions are limited. This article introduces a systematic literature review to label and discuss state-of-the-art studies that deal with SA for hotel reviews.

CCS Concepts: • **Computing methodologies → Natural language processing**; • **Information systems → Information retrieval**; • **Mathematics of computing**;

Additional Key Words and Phrases: Hospitality industry, Sentiment Analysis, machine learning, text mining, hotel reviews

1 INTRODUCTION

The Internet has exploded in the past 15 years, and many tools have sprung up to help people share information and get opinions [140]. Indeed, each Internet user can express different opinions (positive, negative, and neutral), representing valuable information for their decision making [115]. Following each stay at a hotel, a significant portion of hotel customers share their experiences to express what they liked or disliked [134]. Thus, many customers consult online user-generated reviews to decide on hotels. As a general rule, they book the hotels especially after reading reviews of the hotels on websites. In addition, many travelers' beliefs, choices, and decisions are mostly based on what other users think and say. However, hotel service and customer satisfaction indexes are important for managers in the hospitality industry to use customer reviews to learn more about what customers want [107]. They can profit from these reviews by knowing how consumers feel

Authors' addresses: A. Ameur, Polytechnic School of Tunisia & LIPAH, LR111417, Tunis El Manar University, Faculty of Sciences of Tunis, Tunisia; email: asma.ameur@ept.rnu.tn; S. Hamdi, LIPAH, LR111417, Tunis El Manar University, Faculty of Sciences of Tunis, Tunisia; email: sana.hamdi@fst.utm.tn; S. Ben Yahia, Department of Software Science, Tallinn University of Technology, Akadeemia Tee 15a, 12618, Tallinn, Estonia; email: sadok.ben@taltech.ee.

about their services and those of their competitors, which will improve the quality of their hotel products and marketing strategies [74, 150]. In this way, online consumer reviews become a vital business value for the hotel's performance [146]. This enables the concept of "e-reputation," also known as "digital reputation," to be introduced. In a social network, e-reputation becomes a pivotal issue [80]. Thus, it is a compelling problem for the hospitality industry and affects how likely customers are to stay at the same hotel. Well-controlled e-reputation can become a reliable and efficient way to boost the hotel business and construct an effective strategy to manage and improve its communication. Otherwise, the consequences of neglecting its popularity on the Internet for too long can be dramatic. Consequently, the hospitality industry needs **Sentiment Analysis (SA)** applications to automatically identify and analyze this volume of opinions available in textual data. Reciprocally, the hotel domain represents an ideal application for SA, as hotel review websites, such as Tripadvisor, provide many consumer opinions and behavioral data containing rich word-of-mouth information for data analytics [62].

Hence, with increasing hotel marketing efforts and techniques to adopt efficient e-commerce personalization and create smart offers, the significance and popularity of SA research in hospitality are expected to rise further. Given the lack of comprehensive review studies in this area, it is appropriate to assess state-of-the-art SA research in hospitality and provide a basis for future research directions in this area. This article presents a **Systematic Literature Review (SLR)** that scrutinizes 90 papers on SA in the hospitality industry to provide a detailed review with helpful guidelines in this domain. In addition, this SLR targets the identification of gaps and suggests future research directions to give hospitality researchers looking into this topic a place to start. The broad strokes of this SLR are as follows:

- We provide a well-structured work on SA for the hospitality industry by the bias of a robust methodology [76].
- We highlight the key issues and challenges with existing research in SA for the hospitality industry in terms of dataset sources, data preprocessing techniques, feature representation, sentiment classification levels, methods, protocols of validation, and much more to point out key areas for future research in this field.

This article is organized as follows. Section 2 sketches some key concepts. Section 3 presents the research methodology used. Section 4 discusses the review report to answer the different **Research Questions (RQs)**. Then, the challenges of SA in the hospitality field are discussed in Section 5. Finally, Section 6 presents the conclusion of this survey.

2 SA OVERVIEW

People are expressing their moods and emotions in an extensive amount of web data. That is why it is of utmost importance to analyze their reviews. In this section, we cover the fundamental concepts of SA. Then, we sketch out the context dependency problem and explicitly introduce SA's approach to hospitality. Finally, we present a literature review of the survey papers covering SA in hospitality.

2.1 SA Definition

SA, also known as "opinion mining" or "opinion analysis," took off at the beginning of the year 2000 and became one of the most ongoing research areas in **Natural Language Processing (NLP)** [85, 141]. SA is the science of looking at textual data to determine how people feel. It represents a sub-domain of Text Mining that analyzes people's opinions from the user's textual reviews [156]. Indeed, SA pays heed to the direction of opinion about an entity or an aspect of an entity, a polarity that can be positive, neutral, or negative. An opinion represents a thought
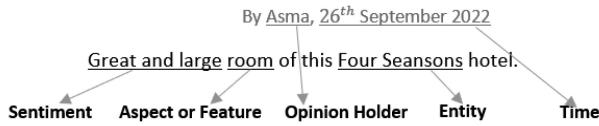
Fig. 1. An example for a sentiment quintuple in the hotel domain.

about something or someone. Its polarity characterizes it as an object—that is, a product, a service, an event, a person, or a company, among others. According to Liu et al. [85], the sentiment is a quintuple (e, a, s, h, t): e is the name of an entity; a is the aspect of the entity e; s is the sentiment on the aspect a knowing that the sentiment s can be a positive, negative, neutral sentiment, or a numeric rating; h is the opinion holder; and t is the timestamp when the opinion is expressed. The quintuple "(Four Seasons, Room, positive, Asma, Sept-26-2022)" presents the sentiment components extracted from the sample review illustrated in Figure 1. It is worth mentioning that not all items in the quintuple are mandatory. For example, the timestamp can be ignored if a fixed time window is considered. Therefore, we derived the tasks of SA from this sentiment quintuple.

SA collects people's opinions and impressions to determine how customers feel about various topics, products, subjects, and services [141]. In general, this is based on the sentiments the customer communicates through a piece of text online toward one object. SA is applied using two basic approaches: lexicon based and **Machine Learning (ML)** [94]. Generally, the objectives of the SA research are as follows [60]: subjectivity detection to find whether the text is opinionated, sentiment classification to predict the text polarity, aspect-based sentiment summarization to glance at the sentiment using the form of scores, and product feature extraction from the review.

2.2 Domain Dependency for SA Tasks

The domain dependency in SA represents a challenging problem to be tackled [133]. This is considered a significant issue in sentiment classification performance as the orientation of words varies by content domain [38]. In addition, online reviews vary by domain in terms of features, sentiments, and relationships between these features that impact the model's training performance. Remarkably, the review aspects present domain dependence that differs from context to context.

For instance, in the hospitality industry, we are interested in "rooms," "Food_Drinks", "service", "location", and so forth. However, for the laptop domain, the interesting aspects are "CPU", "Battery", "Display", "keyboard", and so on. Moreover, depending on the domain, a positive or negative word can have opposite orientations [60]. For example, the meaning of the adjective "small" in mobile and hospitality is different. This adjective presents a favorable opinion of the mobile, saying, "The phone size is small."Contrarily, it shows a negative opinion, saying, "The hotel's room is small." Similarly, "funny movie" is positive in the movie domain, but "funny taste" is negative in the food domain [135]. As a further example, the word "frozen" in software engineering generally presents a negative sentiment. However, it can also elicit a positive response in air-conditioning and refrigeration [60].

2.3 SA Using Hotel Reviews

SA uses hotel reviews to determine whether a portion of guest reviews is positive, neutral, or negative. SA may identify the hotel consumer's attitude toward hotel services. Managers may use SA to collect previous consumer experiences and their findings in service improvement to better contribute to consumer satisfaction and service recovery [80]. SA in the hospitality industry helps hoteliers understand consumer satisfaction regarding their offerings and provides a better understanding of how they differ from their competitors. SA can also help new customers learn about hotel services before purchasing.

2.4 Related Review Studies on SA in Hospitality

Several literature survey papers were proposed to summarize research trends in SA(e.g., [3, 55, 86, 94, 100, 119, 141, 148, 153, 156]). Those studies overview the SA task without going into great detail about any area. However, SA is a field of research that depends on the area's criteria. Indeed, the efficiency of an SA process might fail when applied to a new domain. For example, a model trained on restaurant data cannot be applied to hotel data, especially for the different sub-tasks of aspect extraction [100]. In addition, figuring out the different categories expressed depends a lot on the domain, such as food or the location of a hotel. Thus, it is important to review studies of SA in different research areas by looking at what has already been done to determine where future research should go.

To our knowledge, only four review studies were created on SA in the hospitality domain [12, 70, 89, 95]. Ma et al. [89] present a traditional survey focusing on the SA tools used in hospitality and illustrate their utility with a demonstration study using Tripadvisor data. This survey is presented to hospitality and tourism professionals and researchers studying SA. One of the main problems with this study is that it only briefly looks at the two most important parts of the SA process: extracting features and figuring out how people feel about them. Another limitation is the small number of review articles, which did not exceed 26; all were published before 2017. The work underscored by Alei et al. [12] concludes by outlining future research avenues to further advance SA in tourism as part of a broader Big Data approach. Similarly to the study of Ma et al. [89], this survey reviewed a few articles investigating the SA in the tourism domain. The articles were ancient and published in 2015 or earlier, so none used **Deep Learning (DL)** approaches.

Another recent study by Jain et al. [70] presents an SLR that aims to find the usefulness and applicability of ML techniques for consumer SA on online reviews. First, this study analyzed works using different types of datasets, such as airport, airline, art, and museum reviews, and did not just focus on hotel reviews. Indeed, the number of works using hotel reviews did not exceed 13. Moreover, this study does not discuss the SA process's different parts. Instead, it mostly looks at ML techniques. For instance, it does not give an overview of the different levels at which SA can be performed. Recently, Mehraliyev et al. [95] presented an SLR aiming to analyze the SA literature in hospitality and tourism from methodological and thematic perspectives. Only two analytical techniques (qualitative and quantitative) were used to examine the collected articles. This SLR does not adequately describe the research methodology with simple questions to answer. In addition, it does not look at how ML techniques are used to classify how people feel or how the features of reviews of how people feel are shown. Indeed, authors are content with mentioning the frequency of published papers for most of their analysis topics instead of conducting a critical analysis or even citing these papers. Thus, our current study aims to fill these gaps by doing an SLR to compare, analyze, explore, and understand the attempts and directions of SA research in hospitality. We give a more detailed look at each step in the SA method for hotel reviews, such as preprocessing, feature representation, and classification models. These various aspects are provided for the readers within the RQs.

3 SLR METHODOLOGY

In this section, we present our study based on the methodology of SLR proposed by Kitchenham [76]. The main goal of this research is to accurately study the literature that is already out there, find any flaws in the research that is already out there, and look into the future of SA in hotel reviews. The process of SLR helps determine the different studies available in this domain and answers different RQs. The significant phases of the SLR are highlighted in Figure 2.
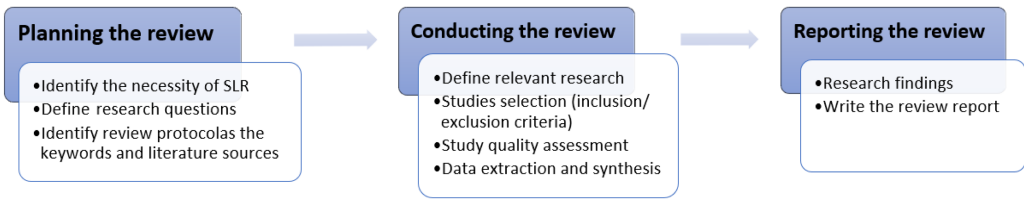
Fig. 2. Review methodology of the SLR stages.

3.1 Review Planning

Next, we describe how this SLR was planned.

3.1.1 Necessity of the SLR. It is necessary to collect the best evidence from the existing literature. The SLR process provides the best techniques to collect and analyze evidence from primary studies. It also addresses the importance of the different methods for each RQ. We used the following search string to see if there is similar literature in this domain: "sentiment analysis" OR "opinion mining" OR "opinion analysis" OR "opinion classification" OR "sentiment classification" AND "hotel" OR "hotel reviews" OR "hospitality" OR "tourism" AND "systematic overview" OR "systematic review" OR "research review" OR "survey." The studies are selected based on their title, abstract, and conclusion. The result of this search depicts a single SLR, introduced in the work of Mehraliyev et al. [95], specific to the SA for hotel reviews. As underscored in Section 2.4, this SLR does not sufficiently describe the research methodology with straightforward questions to answer. Neither does it analyze the use of ML techniques for sentiment classification or the feature representation of customer feedback.

3.1.2 Research Questions. In the following, we present specific RQs and the motivations behind each one:

RQ1: What is the goal of SA in the hospitality industry? This question aims to show the different reasons behind the hotel review analysis and identify its primary objective.
RQ2: What source datasets are used for SA in the hospitality industry? This question tries to present the potential data sources that can be used.
RQ3: What are the sentiment classification levels in hospitality research? This question compares the different levels for SA: document, sentence, and aspect level.
RQ4: What are the different tasks of **Aspect-Based Sentiment Analysis (ABSA)** for hotels? This question describes the distinct steps for the aspect level of SA.
RQ5: Why does hotel SA need textual data preprocessing, and what are its key techniques? question presents the importance of data preprocessing and its techniques, such as stemming and eliminating stop words.
RQ6: What feature representation methods are used for SA in the hospitality industry? In this question, the aim is to highlight the important feature representation methods (classical vs. embedding-based methods).
RQ7: What sentiment classification approaches are used in the hospitality industry? This question aims to define sentiment classification approaches by paying close attention to both lexical- and ML-based approaches (shallow, DL, and hybrid).
RQ8: What validation methods and evaluation tools are used for SA approaches in the hospitality business?

This question shows the validation methods and evaluation metrics that can be used to measure classification quality.

Table 1. Inclusion/Exclusion Criteria for This SLR

Inclusion Criteria	Exclusion Criteria
1. Consider studies that reported on the experiences of SA in the hospitality domain	1. Remove any papers that did not provide useful information to answer the RQs
2. Retain only conference proceedings and scientific journals as potential studies	2. Exclude short papers (posters, summaries of tutorials)
3. Include mainly studies after the year 2010	3. Exclude preliminary papers published by the same author(s)

3.1.3 Protocol Review. The review protocol represents a critical element of any SLR [76]. Here, the adopted review protocol goes into more detail about how this SLR was put together. It lists the search strings, the survey resources, the criteria for inclusion and exclusion, and the rules for judging quality.

Search Terms. To carry out an exhaustive work, we begin by driving the major keywords based on the RQs. The following search query was created by augmenting the keywords with possible synonyms: *sentiment analysis*, *aspect-based sentiment analysis*, *machine learning*, *hotel reviews*, *hotel e-reputation*, *hotel datasets*, *text preprocessing*, and a *classification approach*. Then, the search query was augmented by potential synonyms for these keywords. Moreover, some possible combinations are applied to these terms using Boolean operators ("AND" and "OR") to join them and then create these search strings: (("sentiment analysis" OR "opinion mining" OR "opinion analysis" OR "opinion classification" OR "sentiment classification") AND ("hotel" OR "hotel reviews" OR "hotel e-reputation" OR "hotel dataset") AND ("machine learning" OR "deep learning" OR "aspect-based" OR "text preprocessing" OR "feature representation")). The formed string terms were customized for various databases and provided 778 studies.

Survey Resources. To search the primary papers, we used the following digital libraries and search engines: ACM Digital Library, IEEE Xplorer, Springer Link, Scopus ScienceDirect, arXiv, and Google Scholar.

Study Selection Process. We proceeded with a systematic selection to filter the primary papers by applying the following steps:

- *Step 1*: Review the title to focus on the important SLR research.
- *Step 2*: Examine the studies based on keywords and abstracts to eliminate unrelated studies from our research. If these last details do not give us all of the necessary information, we look at the results and conclusions to ensure that the right papers are chosen. If either of these details is insufficient, the final selection is based on the full text.
- *Step 3*: Remove the duplicated articles obtained from different library databases.
- *Step 4*: To avoid any irrelevant articles, we filter the primary papers by applying the list of our inclusion and exclusion criteria presented in Table 1.
- *Step 5*: Each accepted study in the final set is evaluated regarding the quality assessment rules.
- *Step 6*: Double-check the article's reference list from retained studies to identify additional papers overlooked in the inclusion/exclusion criteria and quality assessment. We mainly consider the papers on SA in hospitality published after 2010, except for a few important studies.

  After each step, the researchers in this study reviewed the papers they had collected and discussed them to ensure that they were worthy of consideration in our SLR, and each disagreement was discussed and resolved.
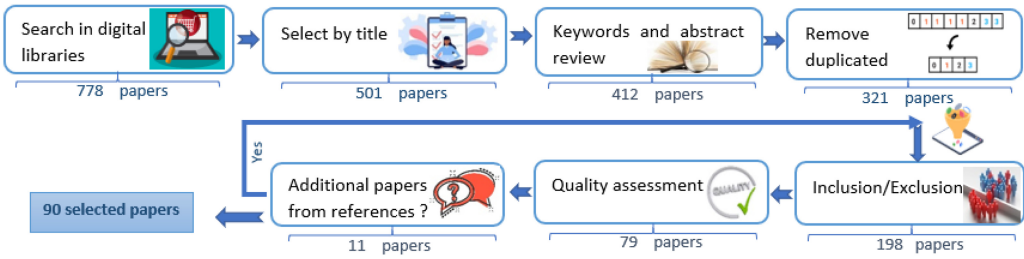
Fig. 3. Search selection protocol for papers relevant to our SLR.

Table 2. Number of the Potential Primary Studies Based on the Data Source

Sources	ACM Digital Library	Science Direct	Springer Link	IEEE Xplorer	arXiv	Google Scholar
Paper Count	58	250	191	183	60	36

Quality Assessment Rules. To evaluate each accepted study in the final set and to gauge the quality of the studies, we consider an adapted quality checklist from the guidelines in the work of Kitchenham [73]. The checklist was applied to assess studies with either Yes = 1 or No = 0. Those with a "yes" answer to the following questions were then selected:

- Is the aim of the study clearly stated?
- Is the paper highly relevant to effectively enriching academia or the industry area?
- Are the methods and data sources for hotel reviews well described?
- Does the research provide a clear methodology with solid findings and clear analysis relevant to the RQs?
- Are there enough metrics to evaluate the output?

3.2 Review Conduction

Following the protocol agreement, we began the selection process and the analysis of primary studies.

3.2.1 Selection of Primary Studies. The process of primary study selection in conducting our SLR is presented in Figure 3. Based on the search string, the initial primary search produced 778 primary papers from different library resources, as shown in Table 2.

3.2.2 Data Extraction. We screened the selected research papers to extract the information needed to report our review. The data extraction was based on different data items; we noted the bibliography (author, title, year, publisher), the article type (journal or conference), the study aims, and the answers to the different RQs. In this step, we were interested in some items to study the distribution of the selected articles in this SLR by years of publication, article types, and publishers. As shown in Figure 4, the distribution of the retained studies demonstrates an increase in SA research in recent years, with a significant expansion after 2014. Moreover, as shown in Figure 5, our statistics reveal that most of the reviewed papers are published by Elsevier, Springer, and IEEE. In addition, most of the retained papers are journal papers. We further observed that more than 60% of the global studies for hotel reviews are in English, as shown in Figure 6.

3.2.3 Data Synthesis. After applying the different selection and filtration steps, we obtained 90 final articles in SA for hotel reviews. Therefore, data extracted from these selected studies was
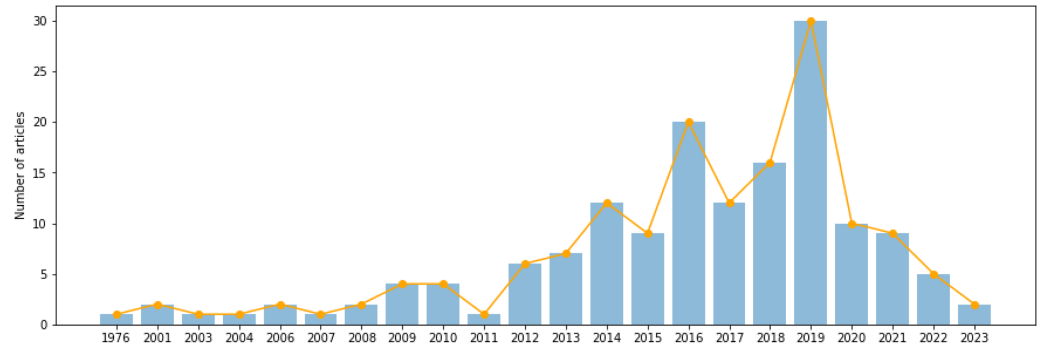
Fig. 4. The distribution of the considered papers by year.
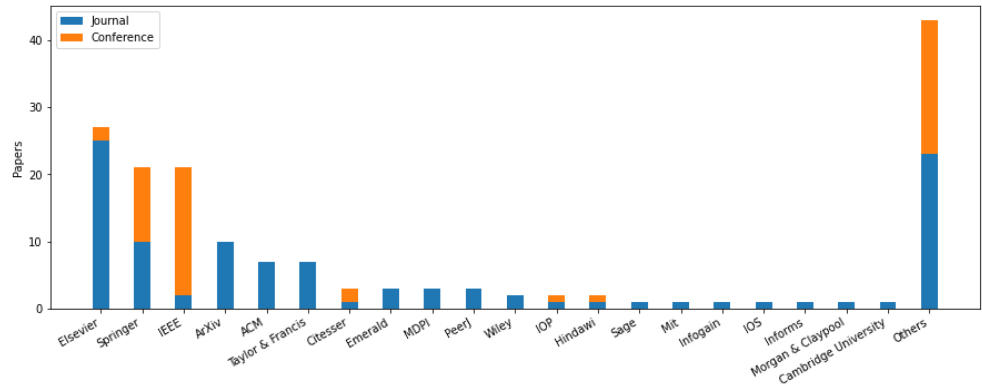
Fig. 5. The distribution of the used papers by publisher.
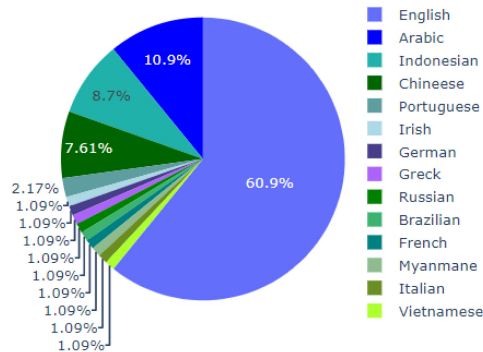
Fig. 6. The distribution of the selected papers by the language of the hotel reviews dataset.

investigated to answer the RQs of this SLR. We followed the guidelines of Kitchenham [76] in synthesizing and reporting results.

4 REPORTING THE REVIEW

This section reports the results of all RQs in accordance with the different steps of the SA process in the hospitality industry. As shown in Figure 7, after applying preprocessing techniques to clean
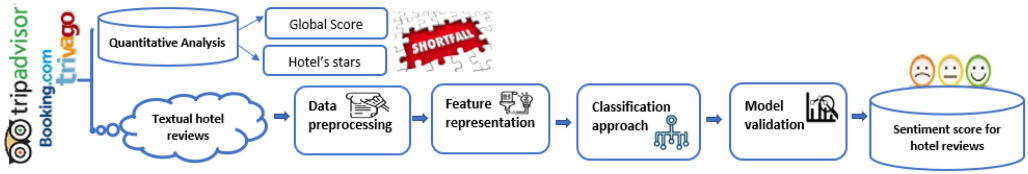
Fig. 7. Framework of SA for the hotel reviews.

Table 3. Goal of SA Tasks in the Hospitality Industry According to the Based Score Method

Method	Description	References
Quantitative analysis	Based on quantitative statistical analysis: global score, number of hotel stars, number of reviews, personal interviews, questionnaires, etc.	[62, 65, 120, 139, 150]
Analysis of hotel reviews	Based on analyzing the textual reviews of Internet users to investigate the user's satisfaction	[1, 2, 7–11, 13, 15, 17, 18, 20, 22, 24–26, 30, 31, 33–35, 37, 40–43, 45–47, 49, 50, 52, 56, 57, 59, 61, 63, 64, 67, 69, 71, 72, 74, 75, 77, 78, 80–83, 87, 88, 91–93, 98, 99, 103, 105–108, 112, 115–118, 121, 123, 125, 128, 130, 131, 134, 137, 142, 144–147, 149, 151, 152, 154, 157]

and tokenize the textual hotel reviews extracted from various opinion websites such as Tripadvisor, a feature representation method was applied to characterize the review. Features extracted from the hotel reviews were then fed to a classifier to train it. Once validated, the trained classifier was finally used to determine the polarity of new reviews. In the following, all results for each RQ are reported separately.

4.1 RQ1: What Is the Goal of SA in the Hospitality Industry?

The goal of opinion mining for the hospitality industry can vary considerably, as shown in Table 3. Sometimes we can only base hotel customer satisfaction on a global overview considering a statistical and quantitative analysis, as presented in some works [62, 65, 120, 139, 150]. For instance, how many users or reviews rate a hotel as positive or negative? For these cases, global hotel polarity classification is sufficient without having to go into details. However, for many reasons, most cases use a more fine-grained approach based on textual review analysis.

Reason 1: The Existing Overall Rating System Is Meaningless and Different from the Written Reviews. Observing the evaluations of hotel users, we can see that many people provide high scores, but their reviews present several problems with the hotel [37]. Thus, these scores cannot be considered reliable. Therefore, the reviews give better insights into the hotel's e-reputation [7]. The score provided on the opinion website cannot sufficiently express the actual satisfaction. The reviews must provide clear information on what guests consider positive or negative about the hotel's services. According to Kasper and Vela [72], we present the review examples for the Ibis and Aston hotels. A user gave a score of 4 stars to the Ibis hotel via this expressed review: "Bad service and disappointing facilities." In contrast, a guest of the Aston hotel provided a score of 3 stars using these reviews: "I like to stay here a lot. The wall and lobby decoration is full of Indonesian reliefs and carving." As such, comments are more significant than the current overall rating on the online websites for expressing customer satisfaction. Details in online reviews clearly explain the consumption experience and perception of the consumer [147]. This can improve the hotel's income and provide insights into service improvement.
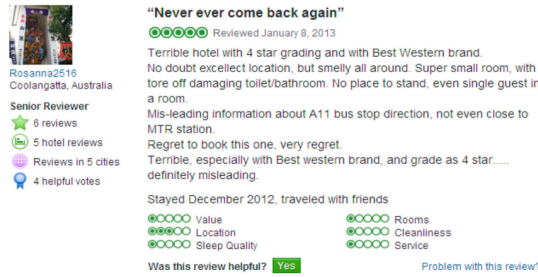
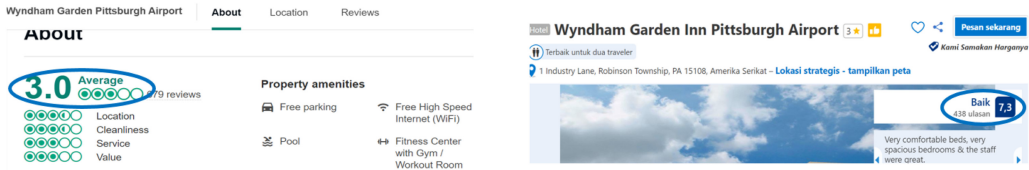Fig. 8. An example of a suspicious rating for a hotel from Tripadvisor [120].

Fig. 9. An example of a comparison of Tripadvisor and Booking.com ratings for the same hotel during the same period.

Reason 2: Contradiction between the Overall and the Specific Rating on the Hotel's Opinion Websites. The rating provided by hotel review websites is the most skewed and cannot be used to fairly assess a hotel's merits [46]. Indeed, the correlation between the overall review rating and specific ratings is very low [109]. This explains why the overall numerical ratings may be better customer satisfaction indicators. To better illustrate, Figure 8 provides further evidence about the contradiction between the overall rating and the specific rating, with a 3.67 gap: the overall rating is 5, and the specific ratings are at an average of 1.33 [120]. This motivation underscores the importance of the hotel's online reviews for customer satisfaction.

Reason 3: Difference between Global Ratings on Multiple Hotel Opinion Websites. To get an idea of a hotel's customer satisfaction, each user can consult one of multiple opinion websites. Figure 9 presents two ratings from Tripadvisor and Booking.com websites for the same hotel and on the same day. Notwithstanding, we notice that we do not have the same global rate, although it concerns the same hotel. Indeed, we can interpret this in different ways. First, one or both ratings differ from the reviews' content, as shown in Reason 1; second, global ratings can be suspicious. For example, we can purposely give unrealistic ratings as manipulation or fake news to increase or decrease sales. Another common reason for the difference between the two ratings is that users who are not serious act in a carefree way. To avoid making misleading decisions, the use of SA can be a helpful solution for managers to improve their hotels and for guests to find their preferred hotels.

In the next section, we pay close attention to the source dataets used in the hospitality field.

4.2 RQ2: What Source Datasets Are Used for SA in the Hospitality Industry?

On the Internet, there are diverse sources that can help guess hotels' customer satisfaction. On the one hand, there are reviews found in hotel or catalog descriptions on the hotels' home pages [72]. On the other hand, more actual and detailed reviews and less biased marketing are available on the Internet. These customers' opinions are more influential, especially when posted on opinion sites. Indeed, multiple review websites for hotels exist. Tripadvisor[1] is one of the most well known

---

[1]https://en.wikipedia.org/wiki/Tripadvisor

Table 4. Opinion Websites Used for Analyzing the User's Satisfaction Based on Hotel Reviews

Website	References
Tripadvisor	[1, 2, 7–11, 13, 15, 17, 20, 22, 24–26, 30, 31, 37, 41, 43, 45, 49, 50, 57, 59, 64, 78, 80, 81, 83, 87, 88, 91–93, 106, 107, 112, 115–118, 120, 128, 130, 131, 134, 137, 143, 146, 149, 151, 152, 154, 157]
Booking.com	[1, 33, 40, 69, 74, 77, 118, 147]
Traveloka	[67, 75, 98, 99, 108]
AiryRooms	[18, 47, 121]
Expedia	[115, 145]
BeerAdvocate	[151, 154, 157]
Agoda	[42, 105]
Datafiniti	[46, 61]
Ctrip	[71, 82, 150]
Trivago	[25]

Table 5. Referenced Dataset Used for SA in the Hospitality Industry

URL of the Dataset for Hotel Reviews	Source	No. of Reviews
http://sifaka.cs.uiuc.edu/ wang296/Data/LARA/TripAdvisor/ [137]	Tripadvisor	108,891
https://github.com/HKUST-KnowComp/DMSC/tree/master/data [151]	Tripadvisor	29,391 [137]
	+BeerAdvocate	+51,020
http://mlg.ucd.ie/datasets/trip/ [143]	Tripadvisor	30,000
https://data.world/datafiniti/hotel-reviews	–	10,000
https://github.com/kavgan/OpinRank/tree/master [49]	Tripadvisor	259,000
https://github.com/txthang/bilstm-crf-lda-hospitality/tree/master/data [131]	Tripadvisor	75,933
https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe	Booking.com	515,000
https://alt.qcri.org/semeval2016/task5/ [106]	Tripadvisor	2,291
http://nemis.isti.cnr.it/~marcheggiani/datasets/ [91]	Tripadvisor	442
https://www.kaggle.com/datasets/ranjitha1/hotel-reviews-city-chennai	Trivago	4,000
https://ieee-dataport.org/documents/hotel-reviews-around-world-sentiment-values-and-review-ratings-different-categoriesfiles [112]	Tripadvisor	69,308

opinion sites and collects more than 463 million visitors per month. This site features more than 8.6 million hotels and other tourist attractions worldwide. Table 4 shows the different opinion websites used in the literature for hotel review analysis. Many studies have exploited the benchmark data to analyze hotel reviews in their research. Thus, a lot of data has been collected and published so that it can be used in the future and put to use in the hospitality industry. Table 5 lists the most widely used and popular datasets for hotel reviews.

4.3 RQ3: What Are the Sentiment Classification Levels in Hospitality Research?

Sentiment classification is the automated process of identifying opinions in reviews and labeling them as positive, negative, or neutral based on the emotions expressed within them. Managing sentiment topics in the hospitality industry assists managers in identifying competitive strengths and weaknesses to provide the target solution [89]. This analysis can occur at different levels: document level, sentence level, or aspect level. Table 6 summarizes SA research works using hotel reviews by level of analysis.

4.3.1 Document Level. The document level only focuses on one topic and predicts if the document polarity expresses positive or negative opinions [85]. This analysis needs to be better for all situations. Indeed, a positive phrase would not unquestionably mean that customers like everything and vice versa—for example, "The hotel was expensive, but the food and the view were great.

Table 6. Summary of Studies Used in the Hotel Domain According to the Level of SA

SA Level	References
Document	[20, 22, 25, 26, 30, 31, 33, 34, 37, 40, 43, 45, 46, 56, 57, 59, 67, 69, 72, 74, 77, 78, 80, 82, 83, 87, 92, 93, 98, 99, 105, 107, 115, 118, 123, 125, 134, 144–147, 149]
Sentence	[24, 41, 42, 91]
Aspect	[1, 2, 7–11, 13, 15, 17, 18, 35, 47, 49, 50, 52, 61, 63, 64, 71, 75, 81, 88, 98, 103, 106, 108, 112, 116, 117, 121, 128, 130, 131, 137, 142, 151, 152, 154, 157]

Table 7. Summary of ABSA Sub-Tasks in the Hospitality Industry

Sub-Task	English	Arabic	Indonesian	Chinese	Myanmar	French
ACD	[7, 13, 52, 61, 64, 75, 108, 112, 117, 121, 130, 131, 142]	[9, 10, 15, 17, 35, 106, 116, 128]	[18, 47]		[63]	[50]
ATE	[117, 121, 131, 142]	[2, 8, 10, 11, 106]	[18, 47]			[50]
SP	[49, 52, 61, 64, 75, 81, 88, 108, 112, 131, 137, 151, 152, 154, 157]	[1, 2, 9–11, 106, 116]		[71]	[63]	[50]

[On the] contrary, the room [was] small and [not] well organized." In the same document, we may express different opinions on multiple aspects of hotel services.

4.3.2 Sentence Level. The sentence level considers the polarity within each sentence, assuming that it contains a single opinion. The sentence level has the limitation of only taking one aspect [119]. Thus, this sentence level cannot be useful for the classification of complex sentences in hotel reviews. Indeed, a sentence can present two opposite opinions about distinct entities—for example, "The hotel was expensive, but the food was great."

Both of the preceding levels can convey a general sentiment. To have a fine-grained analysis, it is essential to use the aspect level.

4.3.3 Aspect Level. For customer satisfaction, only classifying opinions at the document or sentence level is not enough, as we need to identify the opinion of each aspect of the entity to know what users like and dislike [58]. Since it is a fine-grained task [71], ABSA is recommended. Recently, most of the studies focused on the aspect level, as presented in Table 6. The following RQ describes the different sub-tasks for ABSA in the hospitality industry.

4.4   RQ4: What Are the Different Tasks of ABSA for Hotels?

Aspect detection and sentiment polarity issues are not independent [119]. ABSA can classify each opinion according to the aspect categories relevant to each domain. In hospitality, *hotel aspects* influence hotel brand reputation [63]. Besides, based on these aspects, customer satisfaction is supposed to be more accurate than overall user sentiment [75]. Indeed, three sub-tasks of the ABSA task are commonly explored: **Aspect Category Detection (ACD)**, **Aspect Term Extraction (ATE)**, and **Aspect Sentiment Polarity (SP)** [106]. Due to the importance of the ABSA field, several studies are interested in these sub-tasks using different datasets, and Table 7 depicts these sub-tasks according to the language of the hotel reviews:

(1) *ACD:* From a pre-defined set of aspect categories, this task identifies the aspect categories that are indicated implicitly or explicitly in a given review sentence. For instance, in "The hotel was expensive, but the food was great", the aspect categories are "price" and "food." Different methods are used for aspect detection: supervised ML, unsupervised ML, and the hybrid method [119]. Generally, supervised ML approaches perform well to accomplish this sub-task. Note that the performance of these methods depends on the availability of labeled

train data. Moreover, most of these supervised methods require feature engineering to perform well. Thus, many unsupervised methods, such as topic modeling, were proposed to address ACD tasks without the need for labeled data or feature engineering.

In the work of Rana et al. [111], various approaches are reviewed and compared to show the performance of topic modeling in aspect extraction and categorization. We note that the best results for this sub-task are achieved in the work of Khotimah and Sarno [75] using English reviews with an F1 of 84%. For the Arabic language, we mention that the highest result is presented in the work of Ameur et al. [17] with an F1 of 67.30%.

(2) *ATE:* Also known as Opinion Target Extraction, or OTE, ATE aims to identify aspects (attributes) of the target entity in a sentence or a document. This sub-task is to extract a linguistic expression from a text that refers to an aspect of the reviewed entity [58]. The task aims to explicitly extract aspect terms describing features of an entity and opinion terms expressing emotions from user-generated texts [138]—for example, "I liked the location, but not the services." As a result, the location and services are the subjects of targeted expressions of opinion. The study proposed in the work of Septiandri and Sutiono [121] focuses on evaluating transfer learning using pre-trained BERT (Bidirectional Encoder Representation from Transformers) to classify tokens from English hotel reviews in Bahasa Indonesia. Achieving an F-score of 91.4%, this work outperforms related works.

(3) *SP:* This sub-task aims to determine the sentiment polarity of the aspect category. As an example, for the rating of "The hotel was expensive, but the food was great", we would expect a sentiment polarity equal to {price: negative, food: positive}. Likewise, most recent studies for this sub-task used English reviews. We mention that Ray et al. [112] and Khotimah and Sarno [75] highlight the best results for aspect polarity using English reviews, with an accuracy of 92.36% and 94%, respectively. The highest finding using Arabic hotel reviews is presented in the work of Al-Dabet et al. [9] with a validation of 87.31%.

After explaining the different levels of SA, the following section goes over the preprocessing techniques that can be used for cleaning hotel reviews.

4.5 RQ5: Why Does Hotel SA Need Textual Data Preprocessing, and What Are Its Key Techniques?

Data preprocessing aims to remove inconsistent and non-significant data to better extract the information within. It primarily uses traditional methods, like getting rid of things that are not important; changing slang words; and getting rid of URLs, special characters, numbers, parentheses, and punctuation. Text preprocessing presents a potential impact on the performance of the feature extraction [132] and of the text classifier model improvement [32, 132]. Each preprocessing step has an individual influence on the algorithm's accuracy [14] and aims to transform low-quality data into high-quality data. In this respect, the most commonly used techniques for hotel review data are as follows:

- *Tokenization:* Tokenization splits a text into a list of tokens (parts) [67]. The token is a word in a sentence, and it could be a sentence in a paragraph. This is also important for hotel review analysis, as used in several works [47, 74, 88, 98].
- *Normalization:* This technique aims to increase the uniformity of textual data by converting all word forms into one form. Among these techniques used for hotel reviews, we note the following: converting all text to lowercase [59, 74, 75, 93, 108, 131, 137, 147], converting numbers to equivalent words, and correcting misspelled words [74, 75, 107, 108]. Moreover, it includes eliminating empty lines [33, 43], punctuation, parentheses, meaningless

characters, and so forth [10, 15, 17, 69, 93, 108, 131, 137]. Regarding repeated letters, they are quite common in the user's reviews to express their opinions. The repeated letters are, in general, removed, examples of which are '*hungryyyyyyyyyyyy*' and '*huuuuuuuuuuuuungry*' for 'hungry' [126]. Repeated punctuation, with exclamation marks and repeated characters, such as "*haaaaaaappy*," is among the lexical variants that can also intensify the expressed opinions [119]. Thus, the repeated letters can be a challenge in future work because they can underscore the intensity of the opinion polarity. These repeated letters can be converted into a weighted score that conveys emotions. Moreover, it is important to replace contractions with their actual words, which helps with better sentence structuring later. For example, "don't" will be replaced with "do not." This step can be useful to avoid some negative tokens like "can't," "shouldn't," and "don't," which will not be automatically removed via the use of the default stop words list.

- *Stop words removal:* Stop words represent both non-significant terms and common ones in the entire document. They are useless pieces of information that appear with almost equal frequency in all documents. By definition, the stop words group articles, prepositions, and pronouns. They do not have any discrimination power between documents (e.g., "which", "the", "and", "I", "so,""them", "its"). Among the studies that proceed to stop word removal from hotel reviews, we note that there are several (e.g., [13, 33, 37, 50, 59, 61, 69, 74, 75, 93, 98, 108, 137, 147]).

- *Stemming:* This technique aims to transform the word into its root. It comprises deleting the suffix or prefix and adding pre-defined endings to the resulting roots [127]. It is done to keep only the radical part of the words. For example, the stemming of 'en-joying,' 'enjoyed,' and 'enjoys' yields 'enjoy.' This prepossessing step is investigated for hotel reviews in numerous works (e.g., [10, 13, 15, 17, 33, 40, 43, 51, 59, 61, 75, 98, 108, 137, 145, 147]).

- *Lemmatization:* A language dictionary is used to represent the corpus in its original form (canonical or root form). For example "are," "is," and "being" become "be." This is useful for presenting words as descriptors in vector form. It transforms verbs into infinitives and names into singular forms, as in some works (e.g., [93, 107, 131]).

- *Tagging:* To understand the structure of the sentence, the most commonly used method for this issue is based on the **Bag of Words (BoW)** technique. However, the latter cannot detect sentence structure or the syntactic relations between words. That is why it is important to use **Part-of-Speech (POS)** tagging to select the best features and get a fair sentiment classification model. POS tagging is a supervised learning solution for labeling words with their appropriate POS, where a label is the term position or role in the grammatical context [135]. Indeed, POS tagging converts a sentence to a list of tuples, where each tuple has a form (word, tag). The tag explains if the word is a noun, pronoun, adjective, verb, adverb, preposition, conjunction, interjection, and so on. For the sentence "It is a beautiful hotel," we have the output of the POS process as [("It," PRP), ("is," VB), ("a," IN), ("beautiful," JJ), ("hotel," NN), where PRP refers to a pronoun, NN to a noun, VB to a verb, DT to determiner, and JJ to an adjective as presented in the most popular tag set.[2] Multiple studies for SA in the hotel domain use this technique (e.g., [40, 42, 57, 88]). Among the helpful annotation schemes used for data transformation is the format of BIO/IOB (Inside, Outside, Beginning). In addition, the POS tags can be the input of the chunking, which is a critical process to extract information from text.

In the next section, we present an in-depth discussion about how the features are represented.

---

[2]https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

Table 8. Summary of the Features Representations Used in Studies of the Hotel Industry

Feature	References
BoW	[26, 45, 81]
TF-IDF	[17, 20, 33, 43, 45, 56, 61, 92, 112, 117, 118, 123, 147, 152]
N-grams	[17, 22, 42, 43, 45, 72, 78, 106, 123, 130, 152]
Word2Vec based	[2, 8, 9, 11, 45, 50, 67, 69, 71, 77, 98, 99, 105, 112, 131]
Doc2vec	[67, 125]
GloVe	[37, 41, 67, 75, 108, 116, 157]
FastText	[2, 11, 47]
LDA2Vec	[82]
BERT embedding	[1, 15, 112, 149]

4.6 RQ6: What Feature Representation Methods Are Used for SA in the Hospitality Industry?

Feature representation is an essential task for SA [158]. Converting a document to a feature vector is the basic step in any data-driven approach to SA. Using feature vectors to process text more efficiently and decrease dimensional and computational costs is important. Therefore, effective feature selection is important to improve performance and make the learning task accurate. Many studies were interested in effective features using domain expertise and careful engineering. In this SLR, we investigated feature representations for the online reviews of hotels to summarize the sentiment degree (positive or negative), as shown in the work of Ma et al. [89]. There are many feature representation methods, such as BoW, **Term Frequency (TF)**, **Term Frequency–Inverse Term Frequency (TF-IDF)**, and word embedding (static and contextualized) [6], to name but a few. Table 8 summarizes most of the selected studies for hotels based on the techniques used for feature representation. They can be classified as classical and embedding methods, as presented next.

4.6.1 Classical Methods. The classical methods used for feature representation include BoW, TF-IDF, and N-grams.

BoW. BoW is the easiest way to describe a document by checking whether or not it has words that give information. To better understand the BoW process, we present an example of the following two respective sentences [46]: "The best hotel in the USA", and "The Aston hotel is very dirty." A dictionary containing tokens from these two sentences is built as follows: "The," "best," "hotel," "in," "USA," "Aston," "is," "very", and "dirty." We describe each sentence using the following vectors: the first with [1, 1, 1, 1, 0, 0, 0, 0] and the second with [1, 0, 0, 0, 0, 1, 1, 1] [46]. These projections can be represented in the **Document-Term Matrix (DTM)**, where lines represent corpus texts and columns represent descriptors. As the default descriptor, the TF weighting mechanism is a simple digital processing method used for DTM. The formula that computes the weight of the word t in document D is illustrated in Equation (1). Notwithstanding, BoW has some cons, especially when using large corpora, as the vector does not consider the term's importance. In addition, it leads to a high-dimensional space.

$$TF(t, D) = Frequency \ of \ t \ in \ D \tag{1}$$

TF-IDF. This is used to improve text representation and create more sophisticated weights. TF-IDF represents sparse vectors that statistically measure the importance of a word versus the entire corpus. This representation aims to decrease the weight assigned to common words and increase the weight assigned to rare words across all documents. The formula that calculates the weight of words t in a document D is given in Equation (2). As an example, these two types of feature extraction, TF and TF-IDF, are applied for the polarity classification of customer reviews using a

hotel dataset [123]. However, traditional syntactic feature extraction and weighting methods do not consider what the words mean semantically and have high dimensions and sparsity [158]. These latter feature engineering strategies are based on the BoW method, which is information lossless (e.g., the semantics, structure, sequence, and context around nearby words). For example, "We did not find good services, good food, or good presentation" expresses a negative polarity. However, using the BoW technique, we will not detect this because it will process words separately without context.

$$TF - IDF(t, D) = TF(t, D) \times IDF(t) \quad with \quad IDF(t) = \log \frac{Total\ number\ of\ documents\ in\ the\ corpus}{Number\ of\ documents\ containing\ the\ word\ t}$$

(2)

N-grams. These are sequences of words within a given window and classified according to their n value: it is said to be unigrams for $n = 1$, bigrams for $n = 2$, and trigrams for $n = 3$. As an example, consider the sentence "Service is not good." The context and sentiment in this sentence will be better analyzed based on n-grams with $n > 1$ than considering unigram terms. Thus, they can be presented as neighboring sequences of items in a document. N-grams predict the occurrence of a word based on its $n - 1$ previous words' probabilities. The advantage of this technique is its language independence and ability to reflect information about the context. According to the literature review, the n-grams approach helps increase the level of accuracy in SA classification. This technique is widely used for SA problems in the hospitality field and shows promising results. In this context, the study of Laoh et al. [78] produces knowledge about sentiment from the Bali hotel reviews based on the n-gram approach and the **Support Vector Machine (SVM)** model. The findings prove that this technique helps to increase the performance of the hotel review classification and show that bigram results have a higher level of accuracy than the unigram approach. Moreover, a combination of word and character n-grams, as proposed in the work of Siagian and Aritsugi [19], helps to detect deceptive opinions in hotel reviews. This study applies principal component analysis to reduce the feature attributes. The results show that this feature combination overcomes the usual text pattern and provides better results, especially after removing the irrelevant attributes with the principal component analysis technique.

These traditional techniques use similar words and synonyms as independent features and do not grasp the underlying context of the words very well. To solve this challenge of the semantics between words, neural networks make it necessary to explore other techniques for sentiment classification [44]. Recently, DL has shown promising results in SA and outperforms the BoW approach for feature generation [124]. Moreover, DL is popularly used as a powerful technique for feature representation [101]. The embedding methods present a more sophisticated model to capture this loss of information. The following section offers some embedding methods used for hotel review analysis.

4.6.2 Embedding Methods. Known as continuous vector representations of words, they aim to transform each word into a low-dimensional and continuous vector. They encode word embedding as a distributed representation of words and their relative meanings as a dense vector representation. Contrary to the TF-IDF sparse matrix, the word embedding vector captures the context in which words appear. Various word embedding models are used for analysis of hotel reviews, such as Word2Vec, Global Vector (GloVe), Doc2vec, FastText, and LDA2Vec, to name but a few. To initialize the recent word embedding, fine-tuning is used to grab syntactical and semantic information. Generally, the feature representation performance depends on the used data and the field considered. Because extracting features influences classification results, one of the SA challenges is better investigating word embedding [50].

Word2Vec. Word2Vec is one of the most popular techniques, created by Mikolov et al. [96] in 2013, to associate words using points in space. It establishes word meanings and relationships between words. Similar words will be grouped to form a sentiment information class. They base Word2Vec on two key concepts [105]. First, words with similar meanings will be closer together in the space presentation. Second, the vector between the points of two words presents the word relationship. They presented Word2Vec with two different architectures: **Continuous Bag of Words (CBOW)** and skip-gram [96]. Word2Vec is used to detect similar-meaning words to be regrouped together, which is validated as an effective concept [105]. This word embedding technique helps to discover words sharing the same semantics automatically. The CBOW method predicts the target word by the words around it, and the skip-gram model aims to predict the word window given a single word [77]. Thus, this method is one of the representation techniques used for hotel review analysis in research studies. Word2Vec is used in the work of Jiang et al. [71] to capture each word's local features in the hotel reviews corpus for a supervised classification approach. Using European hotel reviews, continuous skip-gram provides good accuracy in the sentiment classification (Best, Good, Bad, and Worst), as the latter works well even for infrequent words [77]. Moreover, skip-gram shows the highest accuracy using Indonesian hotel reviews because it provides a good presentation for the rare words [99].

Doc2vec. To resolve the word order problem, Doc2vec (paragraph vectors) is proposed. It is an unsupervised document vector algorithm that can detect relationships among words and understand the semantics of the text [121]. Indeed, it builds on Word2Vec's learning goal by connecting labels between documents and words rather than words with other words. Two architectures characterize Doc2vec: the Distributed Memory Model of Paragraph Vectors (PV-DM) and the Distributed Bag-of-Words version of Paragraph Vectors (PVDBOW). As an example, this feature representation technique is used in the work of Shuai et al. [125] to analyze hotel reviews.

GloVe. GloVe was developed by Pennington et al. [104]. GloVe is an unsupervised learning model that produces word representations in a clear space where the distance between words is characterized by semantic similarity. The GloVe model can also generate dense word vectors as Word2Vec, but it does so differently. The GloVe model combines the advantages of the skip-gram method and matrix factorization (co-occurrence matrix). The significant advantage of GloVe representation is its ability to model vectors from global corpus terms. However, it is essential to mention that both Word2Vec and GloVe fail to provide the vector representation for any existing words in the model dictionary.

As an example, the GloVe representation is also used for hotel review analysis because it can model vectors representing terms from the global corpus [75]. This latter study shows that GloVe works on the sentiment classification of the hotel aspect level. In the work of Imaduddin et al. [67], hotel review data is used to compare the performance of distinct word embeddings (Word2Vec (skip-gram and CBOW), Doc2vec, and GloVe). As a result, GloVe representation has the highest accuracy, equal to 95.52%.

FastText. FastText is an extension of the Word2Vec model [29]. The main difference between Word2Vec and FastText is that Word2Vec learns from complete words, whereas FastText can learn vectors from the characters of each word's n-grams. We consider this model a BoW model with a sliding window. FastText can capture shorter words and understand suffixes and prefixes. It works well for rare words and guesses with known words in the sentence for missing words. It is used for the feature representation of the hotel reviews to extract the aspects and opinion terms [47, 121].

LDA2Vec. This is a topic model that aims to detect the key topics of a review and enrich the representation of the word vector in terms of the context in which it appears [82]. This

Table 9. Comparison of the Feature Representation Techniques Based on the Word Embedding Used for Analyzing Hotel Reviews

| Ref.	Feature Representation	Best Representation	F-score
[45]	Word embedding (Word2Vec)/BoW/TF-IDF	TF-IDF	85%
[47]	FastText: general embedding/domain embedding/double embedding (general+domain)	Double embedding	91.4% for aspects/ 90% for terms
[67]	CBOW, skip-gram, Doc2vec, and GloVe	GloVe	95.52%
[99]	Word2Vec (CBOW vs. skip-gram) in terms of evaluation method and vector dimension	Skip-gram (hierarchical softmax and 100 as vector dimension)	92.37%
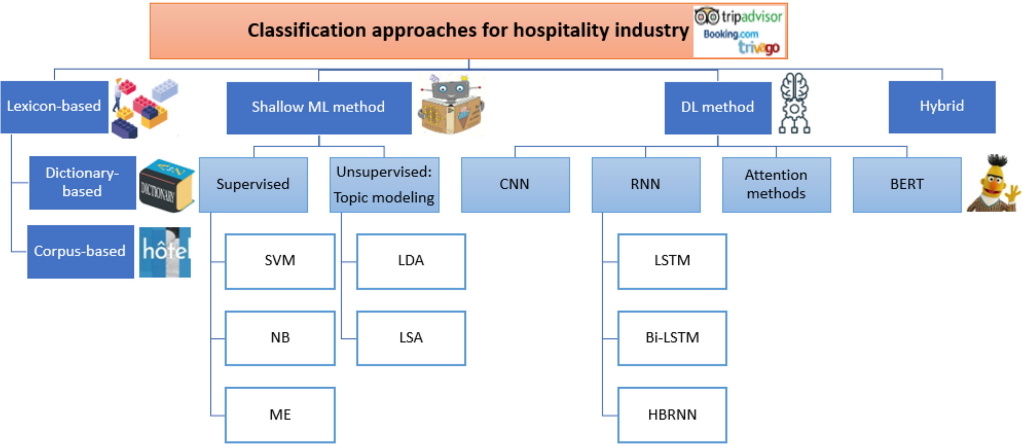
Fig. 10. The different approaches used for SA tasks in the hospitality field.

representation combines Dirichlet theme models for word embedding as it extends skip-gram Word2Vec and **Latent Dirichlet Allocation (LDA)** [97]. Indeed, topic modeling aims to have low-dimensional representations with information gain.

BERT Embedding. BERT was developed by Google and published by Devlin et al. [39]. This model can be used as a text representation to incorporate position and context in a text. A BERT representation is a contextualized word embedding model that considers the word's various contexts.

Many studies are conducted to compare their methods using different feature representations. Table 9 presents these comparative studies and their best results.

After pointing out the multiple ways to build features, we discuss the different models to classify how people feel about hotels in the following RQ.

4.7 RQ7: What Sentiment Classification Approaches Are Used in the Hospitality Industry?

We can broadly categorize the SA classification approaches into two significant groups. The first is a lexicon-based approach, whereas the second is an ML-based approach (shallow ML and DL). The lexical approach does not need labeled data but requires linguistic resources. It is noteworthy that it is difficult for this approach to consider the context. However, the ML-based approach does not require a dictionary. It can deal with ambiguities and is capable of domain adaptation; despite this, it presents a time-consuming task [42]. Figure 10 shows the most commonly used classification approaches for hotel reviews. In Table 10, we examine their advantages and disadvantages. Moreover, Table 11 summarizes the selected studies in the hospitality industry according to these commonly used classification methods.

Table 10. Summary of the Advantages and Disadvantages of the Classification Approach Used in the Hotel Industry Studies

Approach	Advantages	Disadvantages
Dictionary based	Wide coverage in the general domain and are easily transferred from one language to another	It does not detect the specific context of the opinion words
Corpus based	It catches opinion words and their orientations in a specific context	It needs data preprocessing to prepare the corpus and requires a large corpus
Shallow ML	It is simple and does not require a computational resource	It is not very efficient for large datasets
Topic model	It presents low-dimensional representation with information gain	It often gives inaccurate classification results
CNN	It can provide high accuracy with fast training	It is not recommended for long sentences
LSTM	It is able to capture contextual information for long texts	It is unable to catch main parts in the corpus
BRNN	It detects dependency in both directions	Training models are very slow
HBRNN	It is robust, as metrics are aligned with higher resolution in the class structure to improve cost-effectiveness	Similarities agree with hierarchy up to some random noise
Attention	It detects the most useful features and focuses on the important part of the sentence	It uses more parameters, which are time consuming and hard to parallelize
BERT	It learns deeper contextual relations between words	Weights need computational resources

Table 11. Summary of the Classification Approach Used in the Hotel Industry

Classifier	References
Lexical based	[22, 30, 57, 63]
SVM	[10, 17, 24, 26, 33, 34, 42, 45, 78, 83, 92, 105, 106, 118, 123, 125, 152]
NB	[10, 17, 24, 25, 33, 34, 42, 46, 69, 83, 93, 118, 125, 152]
ME	[34, 42, 50]
LDA/LSA/PLSA	[7, 49, 50, 59, 74, 75, 108, 130, 131, 147]
CNN	[2, 8, 9, 37, 41, 71, 116]
LSTM	[9, 67, 69, 71, 75, 98, 108]
BRNN	[2, 8, 11, 47, 69, 82, 131]
HBRNN	[35]
Attention	[8, 9, 47, 71, 82, 151, 157]
BERT	[1, 15, 18, 31, 56, 64, 112, 121, 149]
Hybrid (ML+Lexical)	[61, 118]

4.7.1 Lexicon-Based Approaches. A sentiment lexicon is a list of opinion words and phrases given a positive or negative score to express their sentiment polarity [129]. This approach is based on these opinion words that will be matched with the data to assign a sentiment score to the terms in the dictionary. As this approach does not require training for data classification, lexicon-based learning is considered unsupervised learning [135]. The lexicon-based approaches are split into two sections: dictionary-based approaches and corpus-based ones. For example, the research of Bucur [30] and Gräbner et al. [57] represent two lexical-based studies for English hotel reviews. The two sections of the lexical-based studies are described as follows:

(1) *Dictionary/Thesaurus-based approach:* This type of approach is based on a pre-defined dictionary that provides opinion polarity for the used words. First, the approach aims to find the opinion word in the review text. Second, it considers the semantic relationship between tokens as synonyms, hypernyms, and antonyms from a dictionary [60]. A new iteration adds this term to the seed list for each new word. When there are no new words, the iterative process comes to an end [53]. The total text sentiment score is positive if the document has more positive word lexicons. Otherwise, it is negative [135]. These relations are detected by relying on an external thesaurus [129]. The most common in the hospitality industry are as follows:

— *WordNet*: WordNet is a lexical database of the English language made available to the public by Princeton University. This dictionary returns every sense associated with a word. These groups of senses are called *synsets* [102]. Each one represents a distinct concept using the synonyms and antonyms relationships with the word. Opinion words and their synonyms point in the same direction, and antonyms point in the opposite direction.

— *SentiWordNet*: This dictionary is based on WordNet. In this lexical resource, each synset is defined by three numerical scores to describe how objective, positive, and negative the terms are [21]. This triplet of numerical scores describes the strong terms, knowing that the sum of the affected properties is equal to 1. SentiWordNet is used as an example for hotel review analysis, which calculates the opinion score of each sentence in the reviews based on the summary of component word scores [30]. Moreover, in the work of Wang et al. [137], the opinion for some tested terms among the hotel reviews data is compared with the annotation in SentiWordNet. In addition, this dictionary is used in the work of Chaabani et al. [34] to build a lexicon for SA of different hotel reviews.

— *AFINN*: AFINN is represented by a list of more than 3,300 English words that Finn rup Nielsen (2009 and 2011) manually rates with an integer between −5 (negative) and +5 (positive).[3] This lexicon dictionary is possibly one of the simplest and most widely used lexicons for SA. Using a lexicon-based emoticon and acronyms, Chaabani et al. [34] present a study in the tourism domain to extract sentence polarity.

Dictionary-based approaches are competent for a domain of study, and they work well, especially for a short review. The limitation of this approach is that it cannot find opinion words with domain-specific orientations [135]. Thus, many researchers use a corpus-based method by inducing a sentiment lexicon from text corpora [129].

(2) *Corpus-based approach:* The significant advantage of the corpus-based approach is the ability to find opinion words and their orientations in a specific context [135]. This approach provides sentiment words with a specific context using two sub-categories: a statistical approach via the co-occurrence pattern and a semantic approach via computing the similarity between the terms [53].

4.7.2   ML-Based Approach. In this context, different methods are used to categorize users' opinions of the hospitality industry. Some proposed solutions are based on shallow learning, whereas others are based on DL methods.

Shallow Learning Based Methods. These methods basically include two categories: supervised and unsupervised methods. In supervised learning, the annotators manually label the data and use it to train the algorithm. Thus, the algorithm can classify incoming, unlabeled data based on pre-labeled data. This method outperforms unsupervised methods as it depends on labeled data and includes fewer errors. Unsupervised learning is a lexical-based approach where the data is clustered based on shared features, including word pairings or popular terms. It does not need training data or modeling and instead uses pre-defined lists or dictionaries.

—*Supervised learning*: The different supervised ML models used in SA are generally compared to **Naive Bayes (NB)**, **Maximum Entropy (ME)** (or MaxEnt), and SVM, which are considered references for the ML-based models [94]. Using hotel reviews, Duyen et al. [42] conducted experiments to compare these models with various combinations of features. This study shows that SVM achieved the highest result with 76.8%. Many studies in the hospitality industry are based on these three models to create a hybrid model. For example, Chaabani et al. [34] proposed a hybrid

---

[3]http://www2.imm.dtu.dk/pubdb/pubs/6010-full.html

classification model based on Tunisian hotel reviews shared on Twitter. We express the following in further detail:

(1) *SVM:* Rather than probabilistic classifiers, SVM uses a large margin known as the *Maximum Margin Classifier*. The basic idea behind the training of this supervised algorithm is to find the optimum SVM classifier, called a *hyperplane*, with a larger margin. In this context, most SA research shows that SVM outperforms other shallow ML algorithms in terms of accuracy [135] For example, the SVM model was used to determine the polarity classification of online hotel reviews [123]. This study exploits two types of information: frequency and TF-IDF, where the classification result is better with TF-IDF, with a validated accuracy greater than 87%. The proposed Word2Vec method described in the work of Polpinij et al. [105] is validated through sentiment classification using the SVM algorithm. This approach may enable a more efficient solution for SA because it can help reduce the inherent ambiguity in natural language. Besides, a multi-class classification and regression problem are developed as a review analyzer for overall opinion and aspect opinion in the work of Yu [152]. The latter research shows that SVM presents better results than NB and linear regression. For the challenging task in the OTE, we recommend testing the sequence labeling based on **Conditional Random Fields (CRF)** to compare the results against the SVM performance classifier [10]. Indeed, the CRF model is able to take into account sequential dependencies between segment opinions specific to the same aspect [91].

(2) *NB:* Given the context of class c, NB-supervised classification assumes that all attributes of a given document d are independent. It is a probabilistic classifier that assigns the most probable class $c*$ to the document d. This probabilistic estimation is based on the Bayes rule, as shown in Equation (3). As stated in the work of Martins et al. [93], NB is distinguished by acceptable precision and low computational cost. This technique is straightforward and resists noise and overfitting. The NB classifier is used in several research papers for the sentiment classification of hotel reviews (e.g., [25, 46, 54, 93]). In the work of Ghorpade and Ragha [54], the NB algorithm is used to classify the hotel reviews into positive and negative opinions, precisely for 11 hotels (from Mahabaleshwar City, Maharashtra State, and India). Using the Bayesian model, it is easy to determine the positive and negative opinions for hotel reviews with an accuracy of 96.1%. Using the multinomial NB classifier method, they provided a solution for classifying positive and negative opinions in hotel reviews. In the work of Martins et al. [93], NB is used for hotel review classification. The results show that NB can detect positive and negative polarity. However, the difficulty is in classifying a review as neutral.

$$c^* = argmax_c P(c|d) \qquad where \qquad P(c \,|d) = \frac{P(c) \; P(d \,|c)}{P(d)} \qquad (3)$$

(3) *ME:* This supervised classifier does not make assumptions about the relationships between features. The ME method chooses the model that presents the highest known entropy with the flattest probability distribution [42]. MaxEnt has been proven effective for many NLP applications. Unlike NB, MaxEnt assumes feature independence, which allows for adding features such as bigrams and phrases without the problematic overlapping [34]. Theoretically, ME performs better than NB. Regardless, the latter is more efficient in practice. ME produces accurate results, but the model is challenging to train and may present overfitting [93].

—*Unsupervised learning*: The most popular unsupervised NLP algorithms are the following:

(1) *LDA:* LDA is a probabilistic model used for topic modeling in NLP tasks to find the semantic relations in unstructured documents [28]. This technique is used in different studies of SA

in the hospitality industry, where each document is considered as a mixture of topics mainly used to identify hidden aspects, as underscored in several works [7, 108, 119, 130, 131]. One disadvantage of LDA is that it is designed at the document level and cannot directly intend the aspect level [119]. Indeed, LDA heavily relies on BoW, which makes detected topics global. As a solution, a topic transition called the *Aspect Hidden Markov Model* is used to present a novel probabilistic method for topic segmentation [27]. Besides, an extension of the LDA called *Multi-Grain LDA* is also used to provide global and local topics [119], which can also be tested for hotel reviews.

(2) *LSA:* LSA is also known as latent semantic indexing (or LSI). This unsupervised method is used to provide document clustering or semantic word similarity. It is a linear algebra method that aims to provide the concepts of relationships between the terms of a document using the BoW and a term-document co-occurrence matrix. The most important terms are represented via a decomposition on the DTM by applying singular value decomposition. Indeed, it is helpful to find semantic relations in unstructured documents. In this context, Zhou et al. [158] improved the BoW method by using LSA and LDA to include the semantic analysis and improve the traditional feature representation. **Probabilistic Latent Semantic Analysis (PLSA)**, inspired by LSA, is also used to generate a hidden topic in the term list. As an example, Khotimah and Sarno [75] used the PLSA method for the hospitality industry to yield the hidden topic and underlying semantic structure in the reviews.

In recent years, DL-based techniques have improved SA tasks' accuracy and performed sentiment classification [93]. DL is a highly recommended technique for NLP research, particularly SA for hotel reviews. In the following, we highlight the DL methods used for SA in the hospitality industry:

DL-based methods: DL networks present at least three hidden layers that aim to learn the data representation, where each level transforms the representation into a higher and more abstract representation level [129]. In to a recent overview, Agüero-Torales et al. [4] highlight the advantages of DL approaches for multilingual social media SA. The survey of Abdullah and Ahmet [3] provides diverse DL modes for SA in general, describing their architecture. Several models are used, including **Convolutional Neural Networks (CNN)**, **Recurrent Neural Networks (RNN)**, and the increasing adoption of transformer language models:

—*CNN:* CNNs are the backbone of many modern systems that classify emotions and separate meanings. CNN can generate sentences of fixed-length vectors from sentences of variable length [156]. In this context, in the work of Jiang et al. [71], CNN layers are used to identify the local features for the aspect sentiment classification in a given sentence in Chinese hotel reviews. The CNN model is also used in the work of Ribeiro de Souza et al. [37] for classifying Brazilian hotel reviews. The tests show that CNN achieved considerably good results. Indeed, CNN can be useful for analyzing small sentences [141]. However, it is recommended to use RNN when the size of the sentences grows [37].

—*RNN:* Recently, RNNs have been applied to the NLP issue as SA tasks [68]. These models can process input sequences using their internal memory. A directed graph along the temporal sequence characterizes the connections between nodes. **Long Short-Term Memory (LSTM)**, **Bidirectional Recurrent Neural Networks (BRNN)**, and **Hierarchical Bidirectional Recurrent Neural Networks (HBRNN)** are some of the most common RNN methods used in this field; in addition, all RNNs have feedback loops in the recurrent layers, which let them maintain 'memory' over time:

- *LSTM:* This model is an RNN architecture that can maintain information in memory for a long period of time. It is one of the recommended methods for hotel review classification,

and it presents with an outstanding performance in SA [67]. Priyantina and Sarno [108] used a combination of word embedding and LSTM to conduct the sentiment classification of hotel guest reviews. The performance of this approach to classifying the reviews into five hotel aspects reaches 93% for the ABSA task. The benefit of LSTM is its ability to capture contextual information in long texts. However, it cannot detect the important parts of the corpus. Jiang et al. [71] note that LSTM is not well suited for fine-grained SA because it is presented sequentially and manipulates each context word with the same operation. In addition, LSTM can map out the long-distance dependency, but it is a complex model that requires more memory to train and a lot of time [141]. Extracted convolutional filters for aspect-level sentiment classification are investigated to overcome this limitation. An aspect-based LSTM-CNN attention model is proposed in the work of Jiang et al. [71], exploiting the SemEval-2014 dataset and Chinese hotel reviews. This research combines the ability of LSTM to use the order of the comments and long-range dependencies with the ability of CNN to find local patterns and meanings.

- *BRNN:* These networks can get more information from the outside as they connect two opposing hidden layers on the same output. Like that, the output layer can get information from the past (backward) and future (forward) states simultaneously. They are especially useful when input context is needed. The advantage of BRNN is its capability to detect dependencies in both directions, which provides better results than LSTM, as underscored in the work of Wankhade et al. [141]. In this context, the study by Fernando et al. [47] compares four RNN variations (GRU (Gated Recurrent Unit), LSTM, Bi-GRU (Bidirectional GRU), and Bi-LSTM) based on hotel reviews and using double-embedding methods. Results confirm that Bi-LSTM performs best for both aspect and term extraction sub-tasks. However, this training model is very slow and has high computational costs.

  For hotel reviews analysis, some based combinations are also used, as in the work of Tran et al. [131], who use Bi-LSTM and CRF to identify 10 topics for hotel aspects and their opinions. The results of this study show an F-score of 87%, which confirms the usefulness of the Bi-LSTM–CRF model for the ABSA sub-task in the hospitality field.

- *HBRNN:* Based on similarities and running time, this model is an extension of BRNN, making it more accurate. HBRNN is already used for semantic analysis of hotel reviews in the dataset. This HBRNN model wants to describe how people feel about certain things, saying that the lack of high-quality labeled online reviews presents the biggest problem. This model benefits from the full advantage of RNN for modeling long-term contextual information about temporal sequences in data [35]. As an example, the optimization of this model is achieved using hotel reviews based on the fine-tuning of different parameters [35]. This work compares LSTM, Bi-LSTM, and HBILSTM for sentiment classification, and the results show the superiority of the HBRNN model.

- *Attention Mechanisms and Memory Networks:* This model aims to deduce an attention weight from the lower level to aggregate the weighted vectors in higher-level representation. By modeling the semantic associations, these models can capture the significance of each context word toward a target. That is why these models present an actual improvement in sentiment classification to capture the importance of context words. They concentrate on different parts of a sentence when different aspects are taken as input. Indeed, the attention mechanism presents two main benefits to have better performance [82]: detecting the more useful features and concentrating on the important part of the sentence.

  To analyze the hotel review, a Bi-GRU neural network model was combined with a topic model LDA2vec and an attention mechanism (BiGRULA) is proposed in the work of Li et al. [82]. The authors want to combine LDA2vec and an attention mechanism because not

all words add the same amount to the meaning of a text. LDA2vec is used to discover all key topics in the review corpus. These latter are then used to enrich the word vector representation of words with context. The attention mechanism is used to learn to attribute different weights to the words according to the text's overall meaning. In addition, the approach proposed by Jiang et al. [71] combines LSTM, CNN, and attention models for aspect-level sentiment classification of the hotel reviews. In this study, the LSTM aims to handle long-range dependencies, the CNN identifies local features, and the introduced attention mechanism focuses on the important information of the specific aspect. As underscored in the work of Jiang et al. [71], if we manipulate different aspects, the importance of a word should not be the same as presented in this example: 'good service but the food was dreadful!' The context of 'good' presents more importance than 'dreadful' for the aspect 'service'; however, 'dreadful' is more important for the aspect 'food.'The attention mechanism is also used for multi-aspect sentiment classification of the hotel reviews by proposing an attentive memory network [157]. Furthermore, the targets and contexts need to be learned with special attention. In this way, interactive attention networks are proposed for interactively learning attention in contexts and targets for aspect-level SA. That is why the attention mechanism has been recommended for sentiment classification in recent years.

—*BERT:* BERT is a Transformer-based ML technique for NLP. Like RNNs, Transformers are designed to handle sequential data. However, unlike them, they do not require sequential data to be processed in order. For example, if the input data is a natural language sentence, the Transformer does not need to process its beginning before its end. Because of this, the Transformer is much better at parallelization than RNNs, which means that training times are shorter. As stated in the work of Devlin et al. [39], BERT is a self-supervised method that takes unlabeled data and auto-generates labels by joint conditioning on both left and right contexts in all layers. In recent years, BERT has become a reference ML model because it can handle multiple NLP tasks without human supervision. Thus, using the self-attention model, BERT can detect well the dependencies and important parts of the sentence [141]. As BERT is empirically robust, it has been used in several recent studies in the classification of hotel reviews (e.g., [1, 15, 18, 31, 56, 64, 112, 121, 149]). Using large datasets of annotated hotel reviews, BERT showed better performance and much faster training [64]. For example, for an ABSA of Indonesian hotels, a token classification using transfer learning with BERT is evaluated in the work of Septiandri and Sutiono [121]. Indeed, changing the default BERT to a BERT custom model improves the results. On the one hand, BERT transfer learning for Indonesian hotel reviews performs better and achieves up to a 2% difference in F-scores compared to Bi-LSTM. On the other hand, model validation is improved by adding a CRF with auxiliary labels as an output layer to a BERT-based model.

In the next section, we discuss the hybrid approach, which uses both lexicon- and ML-based methods.

4.7.3 Hybrid Approach. The principle of the hybrid approach is to use a sentiment lexicon to detect the polarity that will be used for data training with the ML approach. The main advantage of the hybrid approach is that it provides high accuracy with powerful supervised learning and stability with the lexicon-based approach [135]. Indeed, hybrid model combinations can be presented in parallel or in stages. Researchers know increasingly that the hybrid approach can present accurate results, but it is not frequently used because it has high computational complexity [89]. As an example, the hybrid approach is used for a tourism model analysis, as described in the work of Chaabani et al. [34]. In the latter work, a lexicon method is used to extract the polarity of each sentence, and ML-based approaches (NB, ME, and SVM) are then used for sentiment classification.

Moreover, the hybrid approach is used for the English hotel reviews, as described in the work of Schmunk et al. [118].

Before carrying out the review and classification processes, the model needs to be validated. In the following, we specify the evaluation measures that can be useful in the context of SA in the hospitality industry.

4.8 RQ8: What Validation Methods and Evaluation Tools Are Used for SA Approaches in the Hospitality Business?

In the hospitality industry, testing the model's ability to predict new data that was not used in estimating it is essential to flag problems like overfitting or selection bias and to give an insight on how the model will be generalized to an independent dataset (i.e., an unknown dataset from real reviews). To accomplish this, the authors of the studied articles use either k-fold cross validation or the usual train/test split for the validation techniques.

4.8.1 Train/Test Split. Here, the data is split into two parts to measure the model's performance, usually called the *training set* and the *test set*. The authors used the training data to build their models, and they used the test data to see how well their models work. When the dataset was large enough, as in the examples of the hotel reviews analysis in the work of Li et al. [81] (29, 391 reviews), they used this technique, as in the work of Baccianella et al. [20] (15,763 reviews), as well as in [105] (20,000 reviews).

4.8.2 Cross Validation. In k-fold cross validation, the original dataset is randomly partitioned into k equal-sized sub-samples. A single sub-sample is employed for testing the model, and the remaining $k - 1$ sub-samples are used as training data. Then, we repeat this process k times, where each k sub-sample is used precisely once when testing data. To maintain the model's accuracy, many studies use k-fold when the dataset is too small (e.g., [25, 45, 83, 87, 92, 98, 118, 123, 125, 137]). The study's data size determines the value of k. The k results can then be averaged to yield a single estimate of the statistical significance of the model. Indeed, this technique avoids overfitting and prevents the data from being biased, especially when the dataset is unbalanced. They mainly divided the used datasets into 10 equal-sized folders. For example, in the work of Bhargave et al. [25], each fold contains 180 reviews. Long et al. [87] used eightfold cross validation to have as much data as possible for training and testing sets. For the same reason, the 1,800 hotel reviews used to identify the authentic and fake ones are validated through fivefold cross validation [23].

4.8.3 Performance Measures. The quality of a supervised algorithm is evaluated by comparing the predicted results with the actual values [90]. These useful parameters are based on a confusion matrix. From a classification point of view, terms such as **True Positive (TP)**, **False Positive (FP)**, **True Negative (TN)**, and **False Negative (FN)** are used to compare classes' labels in this matrix: the acronym TP stands for the proportion of actually positive reviews that the model has incorrectly classified as positive, FP for the proportion of actually negative reviews that the model has correctly classified, TN for the proportion of incorrectly assigned negative reviews, and FN for the proportion of actually negative reviews. According to the works reviewed, the most commonly used metrics for hotel reviews are accuracy, precision, recall, and F-score.

Accuracy. Accuracy is a measure related to the total number of data correctly identified concerning the total amount of data, as presented in Equation (4). This metric is used to evaluate different research of hotel reviews analysis (e.g., [1, 2, 9–11, 24, 33, 41, 56, 61, 64, 67, 69, 71, 74, 75, 78, 81, 83, 92, 98, 99, 106, 107, 112, 116, 151, 154, 157]).

$$Accuracy = \frac{TP + TN}{Total\ reviews\ size} \tag{4}$$

Precision (P). Precision is how precise and accurate the model is for those predicted as positive—that is, how many are positive, as defined in Equation (5). Precision is one of the indicators used in the hospitality industry, as presented in the numerous studies (e.g., [10, 35, 37, 42, 47, 50, 61, 63, 64, 69, 75, 83, 87, 92, 93, 105, 108, 112, 123, 125, 131]). Precision is an excellent measure to determine when the costs of the FP are high. In the hotel business, an FP means a bad review has been mistaken for a good one. The hoteliers aim to ensure guest expectations are met, but they cannot improve their services and achieve guest satisfaction since they cannot realize their requests, suggestions, and complaints.

$$P = \frac{TP}{TP + FP} \tag{5}$$

Recall (R). Recall calculates how many of the actual positives the model captures by labeling them as positive (TP), as underscored in Equation (6). Using the same logic, recall will be the metric used to select the best model when a high cost is associated with an FN. For instance, in the hotel guest experience, an FN means a positive review has been identified as a negative. Suppose the recall for the hotel guest experience modeling is low; in that case, a client looking for a good hotel may be misled and miss important information about the recommended hotels. The recall measures are used for hotel review analysis, as we could witness in several works (e.g., [10, 35, 37, 42, 47, 50, 61, 63, 64, 69, 75, 83, 87, 92, 93, 105, 108, 112, 123, 125, 131]).

$$R = \frac{TP}{FN + TP} \tag{6}$$

F-score *(a.k.a. F-measure, or F-gain).* F-score represents the harmonic mean of precision and recall. This synthetic measure is needed to seek a balance between precision and recall, as explained in Equation (7). The F_{score} is used to validate different models for the hotel review classification (e.g., [8–11, 15, 17, 18, 35, 45–47, 50, 63, 64, 69, 71, 75, 87, 105, 106, 112, 116, 123, 125, 131, 149]).

$$F_{score} = \frac{2 \times TP}{2 \times TP + FN + FP} \tag{7}$$

Mean Squared Error (MSE). MSE measures the average of the squares of the errors between the estimated values and the actual ones. It is another validation metric used in the hospitality industry, as is the case of some works [43, 81, 91, 151, 157]. Next, we discuss the challenges found in the hospitality business.

5 CHALLENGES

Customer experience is becoming increasingly important to the hotel industry, which sees it as a way to stay ahead of the competition. However, guest experience modeling shows a complex and challenging task [7]. Concerning the dataset used in the recent study, the lack of SA research in languages other than English and the manipulation of complex sentences are still difficult problems in SA research. Recently, SA, especially the ABSA field, has been an ongoing field of research, and some points need to be looked into more. For that reason, researchers should be dealing with different challenges. In what follows, we explain the main challenges for the SA tasks of hotel reviews.

Specific Domain. Sentimental expressions vary among domains. Indeed, a sentiment classifier trained to classify opinions in one domain may generate lower results in another domain [26]. Considering more contextual information is essential to dealing with the domain specificity in sentiment classification tasks [51]. In the same way, domain dependency is an ongoing area of SA research in the hotel domain. In this context, an approach to building a domain-dependent sentiment dictionary called *SentiDomain* is proposed in the work of Ahmed et al. [5]. The goal is

to learn a set of sentiment clusters that are part of the global representation of the target domain in the sentences. Compared to the state -of- the art, this proposed framework is effective for sentiment polarity detection at sentence and aspect levels. This dictionary has been trained in four domains (restaurant, laptop, phone, and camera), and it could be useful for hotels.

SA Language. As mentioned in Section 3.2.2, English has the greatest number of SA studies, unlike other languages, including Arabic and its dialects, Chinese, and so on. It will be very interesting to make use of other recent resources that have been constructed in different languages.

Data Annotation. On the one hand, it is important to note that ABSA is mostly used with supervised learning, where annotations should be given. SA performance with supervised methods is slightly better than with unsupervised methods [135]. It will improve as more annotated data is used to train the ML model. The variety of data used to train the ML algorithm will help it learn different features to use its database and give the most relevant results in various scenarios. One of the problems with analyzing hotel reviews is getting more annotated data. However, most of the information available for hotel reviews is not labeled.

On the other hand, as expert data annotation is costly and time consuming, some research on sentiment classification deals with unsupervised methods to save human effort. However, using these methods is still challenging since the results obtained from unsupervised learning are not always useful, as we do not have any label or target to confirm their usefulness.

In this way, we can consider using semi-supervised learning for the hotel reviews' analysis, which takes the middle ground and is very useful and interesting. It uses a small amount of labeled data, bolstering a more extensive set of unlabeled data, and overcomes the individual drawbacks of every single approach [50].

Rich Concept-Centric Aspect-Level SA. Among future research, the direction is moving from traditional word-based approaches into semantically rich, concept-centric, aspect-level SA (known as an implicit aspect) [119]. Indeed, combining ML's power with algorithms makes it possible to deal with complex language structures. The current solutions offer a method for detecting aspects, analyzing sentiment, or both. However, research on a systematic classification of implicit-based approaches and their correlation with explicit-based techniques is still missing [90]. The latter review identifies the techniques used for implicit and explicit aspect extraction or a combination of both, providing perspectives on these approaches. We discussed the explicit expressions in the text, but you have to figure out the ones that are unclear from the context. "The service is great" exemplifies an explicit sentiment [119]. However, the sentence "I could not sleep because of the noise," illustrates an implicit sentiment with an implicit target. Thus, this means that someone needs to sleep well, but this has not happened, so the sentence has a negative tone. Unsupervised techniques can be used on unlabeled data to pull out the different explicit or implicit parts of the hotel reviews. We note the statistical analysis for explicit aspect extractions, topic modeling for implicit aspects, and dependency parsing for combined implicit and explicit aspects, as explained in the work of Maitama et al. [90]. Besides, the Joint Multi-grain Topic Sentiment (JMTS), a domain-independent topic model, is proposed to extract the semantic aspects [13]. The latter study was tested using hotel reviews [137] and restaurant reviews. The result outperforms state-of-the-art models in this field. Regarding topic modeling, the W2VLDA approach, an almost unsupervised system for ABSA, is used for hotel review analysis [50]. This system simultaneously does all three of ABSA's main tasks, which is a good sign for the different areas tested. It is important to think about finding hidden parts for future work.

Document-Level Multi-Aspect Sentiment Classification. **Document-Level Multi-Aspect Sentiment Classification (DMSC)** aims to predict the rating of the different aspects in the

review [151]. The study in the work Zeng et al. [154] proposes the method of DMSC as a valuable task to provide recommendations for both users and business owners. In the work of Yin et al. [151], DMSC is used as a machine comprehension issue, introducing a hierarchical iterative attention model. A Hierarchical User Aspect Rating Network (HUARN) is proposed to simultaneously detect user preference and overall ratings [81]. The hierarchical architecture stores information about this method's words, sentences, and documents. The multiple-hop mechanism is used to study attention operations for aspects and documents. In this context, an attentive memory network for DMSC is used to get aspect-aware document representations [157]. The second work aims to show how important keywords are by using how sentences are put together. Thus, the result of predicting how people will rate an aspect is shown using multi-hop attention memory networks to consider information about neighboring aspects. Manipulating two real-world datasets (Tripadvisor and BeerAdvocate), the model achieves state-of-the-art performance [157].

Aspect Aggregation. The majority of recent hotel reviews ignore the relationship between various aspects. However, score aggregation represents a prevalent issue in the SA task. As is underscored in the work of Basiri et al. [24], most of the existing studies in the recent models applied the majority vote, average, or scaling for the score aggregation between the different aspects, which causes a lack of accuracy in the prediction. This topic also includes hotel aspect analysis [119, 157]. In this way, we noticed that several methods for sentiment score aggregation are based on the Dempster-Shafer theory, which was developed by Dempster and extended by Shafer [122]. This theory can quantify the degree to which some evidence supports a particular proposition [24]. In this context, we mention that a hierarchical aggregation method based on Dempster-Shafer theory is proposed in the work of Basiri et al. [24] and was applied to a social dataset, one of which is from Tripadvisor. Wang et al. [137] presented different approaches for sentiment aggregation. This latter research proposes a new data analysis problem called *Latent Aspect Rating Analysis* (LARA) for sentiment aggregation. This probabilistic rating regression looks at each review's latent aspect rating and weight by looking at the whole review. As future work for this method, they suggest finding the possibly unknown parts and not just describing them with a few keywords.

Review Preprocessing. People who use the Internet now write comments in a way that is easy to read, using shortcuts, number words, and abbreviations. Because of this, preprocessing the data is an important first step for hotel review analysis. As an example, we offer some commonly used terms, such as "gud" for "good," "b4" for "before," and "5in" for "fine," [53]. Consequently, it is essential to consider these cases in the preprocessing of hotel reviews. In addition, emoticons are a problem for the preprocessing review, hurting the deduced opinion score.

Another exciting research challenge for SA preprocessing is the pre-defined list of stop words. Most studies on how to classify how someone feels assume that stop words do not matter. However, we mention that removing some stop words, such as "can't," can inverse the sentiment polarity. To improve the classification results, an automated technique for stop word detection is proposed as future work to handle multi-words [50]. Indeed, some negation words, such as "shouldn't", and "can't", and some intensity terms, such as "very", "more", and "most", have an important influence on opinion detection and customer rating. As a solution, removing stop words cannot make an exception for "no," which is required for opinion extraction [37]. A stop word can also wrap up some words characterizing the studied field, as in the case of common words in the hospitality industry like "hotel", and "room" [7]. The city and hotel names can also be considered stop words. To deal with this, it is important to consider the application domain of the reviews.

Comparative Opinions and Conditional Sentences. Regarding the comparative sentences, we compared one aspect to another one [119]. This is an important topic for hotel review analysis

in SA, as it aims to determine which aspect the consumer prefers [48]. In addition, conditional sentences make it hard to determine how people feel about a hotel based on what they say in reviews.

Negation and Irony. Most SA methods have trouble figuring out when someone is being negative or sarcastic, which makes it hard for the hotel review SA to do its job. Negation can reverse the polarity of the sentence, like in this example: *This hotel is not good.* Moreover, in the following example, *I did not find the hotel funny or interesting*, we note that the opinion is negative, but "funny" and "interesting" give it a positive orientation. These issues affect the sentiment of an aspect, sentence, or document by changing the polarity or strength of the expressed sentiment [119].

Most approaches use a simple word distance metric to reduce words affected by a negation keyword [119]. Thus, a good analysis of the negation expressions could also improve the results [50]. Handling the negative expression represents a challenging problem for the SA tasks, as we can express it in different forms [53]. Negation can be explicitly expressed with reverse polarity, as in the case of *I wouldn't say I like the hotel view*, which is not the case in this example, *I do not like the view, but I like the food.* The latter example requires that the polarity of words be changed until another negation term appears. Furthermore, the sequence "not followed by the only" represents a specific case, as it is in this case: *Not only did I like the view but also the room*, where the polarity is not reversed. Examining some Tunisian hotel reviews, the negation, as an explicit feature with unigrams, does not improve the quality of the text [34]. In this way, they recommend combining the unigram and bigram features, as just using bigrams also makes the space very sparse. To handle the negation words, rule-based and fuzzy logic methods are used in the work of Hardeniya and Borikar [60]. Indeed, the Fuzzy Intensity Finder algorithm is used to deduce the weights of individual words by detecting their intensity in the phrase. The latter study is based on the use of SentiWordnet and Smiley's dictionary[4] to score the sentiment of each word on a three-level scale (+ve, -ve, or neutral opinions). Testing this approach in hotel review analysis is also recommended for future work.

Irony detection, a way that people talk to each other, is one of the issues related to the SA task in the hospitality industry [16]. For this problem, a pragmatic context model is proposed in the work of Wallace [136]. The latter investigates different theories of irony discernment using a probabilistic framework to incorporate contextual information about the user. Wallace [136] also presented a conceptual model to deal with valuable expectations for irony detection. In this context, we recommend exploring other datasets and testing these methods for irony/sarcasm detection in the hospitality industry as a future challenge on this topic. The research of Reyes and Rosso [113] identifies the critical components for irony detection using a freely available dataset with ironic reviews from the Amazon website. The initial output of this model results in valuable insights that can also be useful for irony detection in hotel reviews. Consequently, the negation and irony detection problem should be more thoroughly investigated in future work for SA in the hospitality industry.

Fake News. Online reviews have an impact on consumer decisions. Users increasingly use online reviews to share their post-purchase experiences, but the issue is that not all of these reviews are genuine [23]. As in all areas, hotel guests can fall victim to false reviews, whether positive or negative. To get a good idea of how people feel about a hotel, it is essential to determine how honest the reviews are and eliminate the fake ones. However, detecting fake reviews is not a straightforward task. As they emphasize in "Manipulation," hotel owners and their representatives can post some favorable reviews to raise ratings and draw customers. In contrast, negative reviews can be

---

[4]https://www.csh.rit.edu/~kenny/misc/smiley.html

shared by unethical competitors or individuals who have decided to cause problems for a hotel and mislead their customers [66].

Controlling the writer's account is paramount in the fake news adverse impact mitigation action. Therefore, searching for ways to ensure that the person who writes the feedback has visited the hotel is important. Indeed, Wu et al. [143] show the impact of methods for uncovering suspicious hotel reviews, as these dishonest reviews distort the overall popularity ranking. In this context, the study of Zhang et al. [155] investigates the impacts of online reviewers' verbal and non-verbal behavioral features on fake review detection in the hospitality industry. An SLR for fake reviews on online tourism sites is presented in the work of Reyes-Menendez et al. [114]. This review presents a classification of previous studies in this field according to the user analysis unit used for fake news detection: the user profile [110], as well as the user content (textual [36, 79]) and the user behavior [155]. Despite the importance of this issue, only a few studies have paid heed to the existence of manipulated reviews, as in the case of one work [66]. Indeed, the fake reviews are tricky to distinguish [19]. In this way, their detection can be carried out by detecting duplicates and outliers' identification [53]. To select the 'helpful reviews' for a hotel, it is useful to use the author's credibility from the crawled metadata [7]. Moreover, 10 supervised learning algorithms were exploited to distinguish between authentic and fake reviews [23]. This analysis was based on four linguistic clues: comprehensibility, level of detail, writing style, and cognition indicators. As mentioned in the latter paper, comprehensibility refers to readability, word familiarity, and surface-level characteristics. The level of detail refers to how much information there is, how much information there is to perceive, how much information there is about the context, how much variety there is in the vocabulary, and how many function words there are. In this case, the writing-style pattern refers to how emotions, tenses, and emphasis are used. To influence the hotel's reputation, fake news contains more present and future tense than past tense [23]. In terms of how they show what you know, they can be fillers, tentative, causal, insight, motion, or exclusion words. The tested models in the work of Banerjee [23] analyze the gold dataset that contains English reviews in Asia obtained from participants through e-mail instructions (900 authentic and 900 fake reviews). The results for distinguishing between the two terms of the four linguistic clues are thus consistent with other studies that confirm that authentic and fake reviews can be distinguished based on how they are written. Three extra features (i.e., review density, semantics, and emotion) were also used by Li et al. [84] to detect the fake review. Thus, we can test this proposed method in future work for other datasets in the hospitality field. It is also recommended to investigate the semi-supervised learning algorithm for automating the detection of the hotels' fake reviews. In another way, the review titles can provide important information and detect fake news. Indeed, users should pay closer attention to the title rather than just the review description. In addition, as they emphasize in the work of Banerjee et al. [23], the use of exclamation points, nouns, and articles in the titles of the reviews can help distinguish between authentic and fake reviews. Future research on fake news detection will also combine content-based features with the reviewer's behavior, information about the services, and so on to have more information about the hotel.

Recommender Systems. Over the past few years, the number of sites using recommender systems with different models has steadily grown. Review sites have found these systems helpful, as they try to improve the user experience to gain market share and make more money through deals. Hotels are a prime target for this effort, as there is a large number of them in most destinations and a lot of differentiation between them. Suppose that a consumer travels to a new or unfamiliar location or even looks for a new experience. In that case, the Internet is usually the first stop for information to determine his preferences. Customer reviews can change a potential

customer's mind about a restaurant or hotel before the customer even stepped foot on the property. Websites like Tripadvisor, where users can leave reviews for hotels and other tourist activities, are a common way to make decisions because they have ratings for different parts of a business. Thus, it can be hard for users to sort through all of this information and determine what is important.

Researchers have made and studied many recommendation systems, such as regression, collaborative filtering, and matrix factorization. The time a user spends, the number of hits for recommendations, click tracking, and likes or dislikes are all worth considering as essential features for improved recommendation accuracy. In addition, a hotel recommender system can use topic modeling techniques to analyze hotel customer reviews and generate implicit features, as in one work [43]. The diversity between explicit and implicit features enhances the accuracy of the system. Recommendation systems may suggest similar or more diverse items in various situations. At the same time, the most accurate results come from recommending hotels based on how similar users or hotels are. We know this as the "diversity issue," where recommendations are based on overlapping instead of differences. This gives the user a smaller list of hotels to choose from, and niche hotels closely related to their search may be missed. However, the variety of suggestions lets users find out about the value of interesting hotels they might not have found otherwise.

6 CONCLUSION

It is important to point out that SA in the hospitality industry is still developing. This SLR gave an overview of how the hospitality industry uses SA methods to look at online reviews. The answers to the RQs helped us achieve our goal. In this way, our SLR has contributed to building the knowledge of SA for hotel reviews and provided the stakeholders in this domain with information. In addition, it is worth noting that ABSA is about the level recommended for hotel review analysis. Finally, we mention that for future research in the hospitality industry, it is pivotal to explore DL and unsupervised sentiment classification techniques and find more patterns. Combining review data with other features, such as weather and special events, is helpful.

REFERENCES

[1] Mohammed M. Abdelgwad, Taysir Hassan A. Soliman, and Ahmed I. Taloba. 2022. Arabic aspect sentiment polarity classification using BERT. *Journal of Big Data* 9, 1 (2022), 1–15.

[2] Mohammed M. Abdelgwad, Taysir Hassan A. Soliman, Ahmed I. Taloba, and Mohamed Fawzy Farghaly. 2022. Arabic aspect-based sentiment analysis using bidirectional GRU-based models. *Journal of King Saud University–Computer and Information Sciences* 34, 9 (2022), 6652–6662.

[3] Tariq Abdullah and Ahmed Ahmet. 2022. Deep learning in sentiment analysis: Recent architectures. *ACM Computing Surveys* 55, 8 (2022), Article 159, 37 pages.

[4] Marvin M. Agüero-Torales, Jose I. Abreu Salas, and Antonio G. López-Herrera. 2021. Deep learning and multilingual sentiment analysis on social media data: An overview. *Applied Soft Computing* 107 (2021), 107373.

[5] Murtadha Ahmed, Qun Chen, and Zhanhuai Li. 2020. Constructing domain-dependent sentiment dictionary for sentiment analysis. *Neural Computing and Applications* 32, 18 (2020), 14719–14732.

[6] Ravinder Ahuja, Aakarsha Chug, Shruti Kohli, Shaurya Gupta, and Pratyush Ahuja. 2019. The impact of features extraction on the sentiment analysis. *Procedia Computer Science* 152 (2019), 341–348.

[7] Nadeem Akhtar, Nashez Zubair, Abhishek Kumar, and Tameem Ahmad. 2017. Aspect based sentiment oriented summarization of hotel reviews. *Procedia Computer Science* 115 (2017), 563–571.

[8] Saja Al-Dabet, Sara Tedmori, and Mohammad Al-Smadi. 2020. Extracting opinion targets using attention-based neural model. *SN Computer Science* 1, 5 (2020), 1–10.

[9] Saja Al-Dabet, Sara Tedmori, and Mohammad Al-Smadi. 2021. Enhancing Arabic aspect-based sentiment analysis using deep learning models. *Computer Speech & Language* 69 (2021), 101224.

[10] Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yaser Jararweh, and Omar Qawasmeh. 2019. Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic and semantic features. *Information Processing & Management* 56, 2 (2019), 308–319.

[11] Mohammad Al-Smadi, Bashar Talafha, Mahmoud Al-Ayyoub, and Yaser Jararweh. 2019. Using long short-term mem-
      ory deep neural networks for aspect-based sentiment analysis of Arabic reviews. *International Journal of Machine
      Learning and Cybernetics* 10 (2019), 2163–2175.
[12] Ali Reza Alaei, Susanne Becken, and Bela Stantic. 2019. Sentiment analysis in tourism: Capitalizing on big data.
      Journal of Travel Research 58, 2 (2019), 175–191.
[13] Md. Hijbul Alam, Woo-Jong Ryu, and SangKeun Lee. 2016. Joint multi-grain topic sentiment: Modeling semantic
      aspects for online reviews. *Information Sciences* 339 (2016), 206–223.
[14] Saqib Alam and Nianmin Yao. 2019. The impact of preprocessing steps on the accuracy of machine learning algo-
      rithms in sentiment analysis. *Computational and Mathematical Organization Theory* 25, 3 (2019), 319–335.
[15] Asma Ameur, Sana Hamdi, and Sadok Ben Yahia. 2023. Arabic aspect category detection for hotel reviews based on
      data augmentation and classifier chains. In *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*.
[16] Asma Ameur, Sana Hamdi, and Sadok Ben Yahia. 2023. Domain adaptation approach for Arabic sarcasm detection in
      hotel reviews based on hybrid learning. In *Proceedings of the 27th International Conference on Knowledge-Based and
      Intelligent Information and Engineering Systems*.
[17] Asma Ameur, Sana Hamdi, and Sadok Ben Yahia. 2023. Multi-label learning for aspect category detection of Arabic
      hotel reviews using AraBERT. In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence
      (ICAART'23)*, Vol. 2. 241–250. https://doi.org/10.5220/0011694800003393
[18] Yosef Ardhito Winatmoko, Ali Akbar Septiandri, and Arie Pratama Sutiono. 2019. Aspect and opinion term extraction
      for hotel reviews using transfer learning and auxiliary labels. *arXiv e-prints arXiv:1909.11879* (2019).
[19] Al Hafiz Akbar Maulana Siagian and Masayoshi Aritsugi. 2017. Combining word and character n-grams for detect-
      ing deceptive opinions. In *Proceedings of the 2017 IEEE 41st Annual Computer Software and Applications Conference
      (COMPSAC'17)*, Vol. 1. IEEE, Los Alamitos, CA, 828–833.
[20] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2009. Multi-facet rating of product reviews. In *Proceedings
      of the European Conference on Information Retrieval*. 461–472.
[21] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for
      sentiment analysis and opinion mining. In *Proceedings of the 7th International Conference on Language Resources and
      Evaluation (LREC'10)*. 2200–2204.
[22] Sayeh Bagherzadeh, Sajjad Shokouhyar, Hamed Jahani, and Marianna Sigala. 2021. A generalizable sentiment anal-
      ysis method for creating a hotel dictionary: Using big data on TripAdvisor hotel reviews. *Journal of Hospitality and
      Tourism Technology* 12, 2 (2021), 210–238.
[23] Snehasish Banerjee, Alton Y. K. Chua, and Jung-Jae Kim. 2015. Using supervised learning to classify authentic and
      fake online reviews. In *Proceedings of the 9th International Conference on Ubiquitous Information Management and
      Communication*. 1–7.
[24] Mohammad Ehsan Basiri, Ahmad Reza Naghsh-Nilchi, and Nasser Ghasem-Aghaee. 2014. Sentiment prediction based
      on Dempster-Shafer theory of evidence. *Mathematical Problems in Engineering* 2014, 3 (2014), 1–13.
[25] P. Sanjay Bhargav, G. Nagarjuna Reddy, R. V. Ravi Chand, K. Pujitha, and Anjali Mathur. 2019. Sentiment analysis
      for hotel rating using machine learning algorithms. *International Journal of Innovative Technology and Exploring
      Engineering* 8, 6 (2019), 1225–1228.
[26] Federica Bisio, Paolo Gastaldo, Chiara Peretti, Rodolfo Zunino, and Erik Cambria. 2013. Data intensive review min-
      ing for sentiment classification across heterogeneous domains. In *Proceedings of the 2013 IEEE/ACM International
      Conference on Advances in Social Networks Analysis and Mining*.
[27] David M. Blei and Pedro J. Moreno. 2001. Topic segmentation with an aspect hidden Markov model. In *Proceedings of
      the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 343–348.
[28] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning
      Research* 3 (Jan. 2003), 993–1022.
[29] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword
      information. *Transactions of the Association for Computational Linguistics* 5 (2017), 135–146.
[30] Cristian Bucur. 2015. Using opinion mining techniques in tourism. *Procedia Economics and Finance* 23 (2015),
      1666–1673.
[31] Luca Cagliero, Moreno La Quatra, and Daniele Apiletti. 2020. From hotel reviews to city similarities: A unified latent-
      space model. *Electronics* 9, 1 (2020), 197.
[32] Jose Camacho-Collados and Mohammad Taher Pilehvar. 2017. On the role of text preprocessing in neural network
      architectures: An evaluation study on text categorization and sentiment analysis. *arXiv preprint arXiv:1707.01780*
      (2017).
[33] Diogo Campos, Rodrigo Rocha Silva, and Jorge Bernardino. 2019. Text mining in hotel reviews: Impact of words
      restriction in text classification. In *Proceedings of the 11th International Joint Conference on Knowledge Discovery,
      Knowledge Engineering, and Knowledge Management (IC3K'19)*. 442–449.

[34] Yasmine Chaabani, Radhia Toujani, and Jalel Akaichi. 2018. Sentiment analysis method for tracking touristics reviews in social media network. In *Proceedings of the International Conference on Intelligent Interactive Multimedia Systems and Services*. 299–310.

[35] Arindam Chaudhuri and Soumya K. Ghosh. 2016. Sentiment analysis of customer reviews using robust hierarchical bidirectional recurrent neural network. In *Artificial Intelligence Perspectives in Intelligent Systems*. Springer, 249–261.

[36] Run Yu Chen, Jin Yi Guo, and Xiao Long Deng. 2014. Detecting fake reviews of hype about restaurants by sentiment analysis. In *Proceedings of the International Conference on Web-Age Information Management*. 22–30.

[37] Joana Gabriela Ribeiro de Souza, Alcione de Paiva Oliveira, Guidson Coelho de Andrade, and Alexandra Moreira. 2018. A deep learning approach for sentiment analysis applied to hotel's reviews. In *Proceedings of the International Conference on Applications of Natural Language to Information Systems*. 48–56.

[38] Shuyuan Deng, Atish P. Sinha, and Huimin Zhao. 2017. Adapting sentiment lexicons to domain-specific social media texts. *Decision Support Systems* 94 (2017), 65–76.

[39] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[40] Taşkın Dirsehan. 2016. Text mining in the hospitality sector to extend the motivation theory. In *Proceedings of the International Marketing Trends Conference*.

[41] Xin Luna Dong and Gerard De Melo. 2018. A helping hand: Transfer learning for deep sentiment analysis. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2524–2534.

[42] Nguyen Thi Duyen, Ngo Xuan Bach, and Tu Minh Phuong. 2014. An empirical study on sentiment analysis for Vietnamese. In *Proceedings of the 2014 International Conference on Advanced Technologies for Communications (ATC'14)*. IEEE, Los Alamitos, CA, 309–314.

[43] Ashkan Ebadi and Adam Krzyzak. 2016. A hybrid multi-criteria hotel recommender system using explicit and implicit feedbacks. *International Journal of Computer and Information Engineering* 10, 8 (2016), 1450–1458.

[44] Mariem ElAbdi, Boutheina Smine, and Sadok Ben Yahia. 2018. DFBICA: A new distributed approach for sentiment analysis of bibliographic citations. In *Proceedings of the 2018 12th International Conference on Research Challenges in Information Science (RCIS'18)*. IEEE, Los Alamitos, CA, 1–6.

[45] Ahmad Naufal Farhan and Masayu Leylia Khodra. 2017. Sentiment-specific word embedding for Indonesian sentiment analysis. In *Proceedings of the 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA'17)*. IEEE, Los Alamitos, CA, 1–5.

[46] Arif Abdurrahman Farisi, Yuliant Sibaroni, and Said Al Faraby. 2019. Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. *Journal of Physics: Conference Series* 1192 (2019), 012024.

[47] Jordhy Fernando, Masayu Leylia Khodra, and Ali Akbar Septiandri. 2019. Aspect and opinion terms extraction using double embeddings and attention mechanism for Indonesian hotel reviews. *arXiv preprint arXiv:1908.04899* (2019).

[48] Murthy Ganapathibhotla and Bing Liu. 2008. Mining opinions in comparative sentences. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING'08)*. 241–248.

[49] Kavita Ganesan and ChengXiang Zhai. 2012. Opinion-based entity ranking. *Information Retrieval* 15, 2 (2012), 116–150.

[50] Aitor García-Pablos, Montse Cuadros, and German Rigau. 2018. W2VLDA: Almost unsupervised system for aspect based sentiment analysis. *Expert Systems with Applications* 91 (2018), 127–137.

[51] A. Geethapriya and S. Valli. 2021. An enhanced approach to map domain-specific words in cross-domain sentiment analysis. *Information Systems Frontiers* 23, 3 (2021), 791–805.

[52] Mohamed Gharzouli, Aimen Khalil Hamama, and Zakaria Khattabi. 2021. Topic-based sentiment analysis of hotel reviews. *Current Issues in Tourism* 25, 9 (2022), 1368–1375.

[53] M. Ghode, S. Bere, M. Kamale, and M. Moholkar. 2014. Sentiment analysis over online product reviews: A survey. *International Journal on Recent and Innovation Trends in Computing and Communication* 2, 1 (2014), 3766–3774.

[54] Tushar Ghorpade and Lata Ragha. 2012. Featured based sentiment classification for hotel reviews using NLP and Bayesian classification. In *Proceedings of the 2012 International Conference on Communication, Information, and Computing Technology (ICCICT'12)*. IEEE, Los Alamitos, CA, 1–5.

[55] Anastasia Giachanou and Fabio Crestani. 2016. Like it or not: A survey of Twitter sentiment analysis methods. *ACM Computing Surveys* 49, 2 (2016), 1–41.

[56] Santiago González-Carvajal and Eduardo C. Garrido-Merchán. 2020. Comparing BERT against traditional machine learning text classification. *arXiv preprint arXiv:2005.13012* (2020).

[57] Dietmar Gräbner, Markus Zanker, Günther Fliedl, and Matthias Fuchs. 2012. Classification of customer reviews based on sentiment analysis. In *Information and Communication Technologies in Tourism 2012*. Springer, 460–470.

[58] Hussam Hamdan. 2016. SentiSys at SemEval-2016 Task 5: Opinion Target Extraction and Sentiment Polarity Detection. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16)*. 350–355.

[59] Hyun Jeong Han, Shawn Mankad, Nagesh Gavirneni, and Rohit Verma. 2016. *What Guests Really Think of Your Hotel: Text Analytics of Online Customer Reviews*. The Center for Hospitality Research, Cornell University.

[60] Tanvi Hardeniya and Dilipkumar A. Borikar. 2016. Dictionary based approach to sentiment analysis—A review. *International Journal of Advanced Engineering, Management and Science* 2, 5 (2016), 317–322.

[61] Alia Karim Abdul Hassan and Ahmed Bahaa Aldeen Abdulwahhab. 2019. Location aspect based sentiment analyzer for hotel recommender system. *Iraqi Journal of Science* 60, 1 (2019), 143–156.

[62] Ángel Herrero, Héctor San Martín, and José M. Hernández. 2015. How online search behavior is influenced by user-generated content on review websites and hotel interactive websites. *International Journal of Contemporary Hospitality Management* 27, 7 (2015), 1573–1597.

[63] Cho Cho Hnin, Naw Naw, and Aung Win. 2018. Aspect level opinion mining for hotel reviews in Myanmar language. In *Proceedings of the 2018 IEEE International Conference on Agents (ICA'18)*. IEEE, Los Alamitos, CA, 132–135.

[64] Mickel Hoang, Oskar Alija Bihorac, and Jacobo Rouces. 2019. Aspect-based sentiment analysis using BERT. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*. 187–196.

[65] Ivanka Avelini Holjevac, Suzana Marković, and Sanja Raspor. 2009. Customer satisfaction measurement in hotel industry: Content analysis study. In *Proceedings of the 4th International Scientific Conference on Planning for the Future Learning from the Past: Contemporary Developments in Tourism, Travel, and Hospitality*.

[66] Nan Hu, Indranil Bose, Noi Sian Koh, and Ling Liu. 2012. Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems* 52, 3 (2012), 674–684.

[67] Helmi Imaduddin, Widyawan, and Silmi Fauziati. 2019. Word embedding comparison for Indonesian language sentiment analysis. In *Proceedings of the 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT'19)*. IEEE, Los Alamitos, CA, 426–430.

[68] Ozan Irsoy and Claire Cardie. 2014. Opinion mining with deep recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP'14)*. 720–728.

[69] Abid Ishaq, Muhammad Umer, Muhammad Faheem Mushtaq, Carlo Medaglia, Hafeez Ur Rehman Siddiqui, Arif Mehmood, and Gyu Sang Choi. 2021. Extensive hotel reviews classification using long short term memory. *Journal of Ambient Intelligence and Humanized Computing* 12 (2021), 9375–9385.

[70] Praphula Kumar Jain, Rajendra Pamula, and Gautam Srivastava. 2021. A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review* 41 (2021), 100413.

[71] Ming Jiang, Wen Zhang, Min Zhang, Jianping Wu, and Tao Wen. 2019. An LSTM-CNN attention approach for aspect-level sentiment classification. *Journal of Computational Methods in Sciences and Engineering* 19, 4 (2019), 859–868.

[72] Walter Kasper and Mihaela Vela. 2011. Sentiment analysis for hotel reviews. In *Proceedings of the Computational Linguistics-Applications Conference*, Vol. 231527. 45–52.

[73] B. Kitchenham and S. Charters. 2007. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. EBSE Technical Report, Version 2.3. EBSE.

[74] Dewi Ayu Khusnul Khotimah and Riyanarto Sarno. 2018. Sentiment detection of comment titles in Booking.com using probabilistic latent semantic analysis. In *Proceedings of the 2018 6th International Conference on Information and Communication Technology (ICoICT'18)*. IEEE, Los Alamitos, CA, 514–519.

[75] Dewi Ayu Khusnul Khotimah and Riyanarto Sarno. 2019. Sentiment analysis of hotel aspect using probabilistic latent semantic analysis, word embedding and LSTM. *International Journal of Intelligent Engineering and Systems* 12, 4 (2019), 275–290.

[76] Barbara Kitchenham. 2004. Procedures for performing systematic reviews. *Keele, UK, Keele University* 33, 2004 (2004), 1–26.

[77] Shreyas Rajesh Labhsetwar. 2019. Sentiment analysis of customer satisfaction using deep learning. *Res J Comput Sci (IRJCS)* 6 (2019), 709–715.

[78] Enrico Laoh, Isti Surjandari, and Nadhila Idzni Prabaningtyas. 2019. Enhancing hospitality sentiment reviews analysis performance using SVM N-grams method. In *Proceedings of the 2019 16th International Conference on Service Systems and Service Management (ICSSSM)*. IEEE, Los Alamitos, CA, 1–5.

[79] Theodoros Lappas, Gaurav Sabnis, and Georgios Valkanas. 2016. The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research* 27, 4 (2016), 940–961.

[80] Huiying Li, Qiang Ye, and Rob Law. 2013. Determinants of customer satisfaction in the hotel industry: An application of online review analysis. *Asia Pacific Journal of Tourism Research* 18, 7 (2013), 784–802.

[81] Junjie Li, Haitong Yang, and Chengqing Zong. 2018. Document-level multi-aspect sentiment classification by jointly modeling users, aspects, and overall ratings. In *Proceedings of the 27th International Conference on Computational Linguistics*. 925–936.

[82] Qin Li, Shaobo Li, Jie Hu, Sen Zhang, and Jianjun Hu. 2018. Tourism review sentiment classification using a bidirectional recurrent neural network with an attention mechanism and topic-enriched word vectors. *Sustainability* 10, 9 (2018), 3313.

[83] Xiaoyu Li and Cai Liu. 2020. Comparison of machine learning models for sentimental analysis of hotel reviews. *IOP Conference Series: Materials Science and Engineering* 806 (2020), 012029.

[84] Yuejun Li, Xiao Feng, and Shuwu Zhang. 2016. Detecting fake reviews utilizing semantic and emotion model. In *Proceedings of the 2016 3rd International Conference on Information Science and Control Engineering (ICISCE'16)*. IEEE, Los Alamitos, CA, 317–320.

[85] Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Springer.

[86] Bing Liu and Lei Zhang. 2012. A survey of opinion mining and sentiment analysis. In *Mining Text Data*. Springer, 415–463.

[87] Chong Long, Jie Zhang, and Xiaoyan Zhu. 2010. A review selection approach for accurate feature rating estimation. In *COLING 2010: Posters*. 766–774.

[88] Jian Ming Luo, Huy Quan Vu, Gang Li, and Rob Law. 2021. Understanding service attributes of robot hotels: A sentiment analysis of customer online reviews. *International Journal of Hospitality Management* 98 (2021), 103032.

[89] Emily Ma, Mingming Cheng, and Aaron Hsiao. 2018. Sentiment analysis—A review and agenda for future research in hospitality contexts. *International Journal of Contemporary Hospitality Management* 30, 11 (2018), 3287–3308.

[90] Jaafar Zubairu Maitama, Norisma Idris, Asad Abdi, Liyana Shuib, and Rosmadi Fauzi. 2020. A systematic review on implicit and explicit aspect extraction in sentiment analysis. *IEEE Access* 8 (2020), 194166–194191.

[91] Diego Marcheggiani, Oscar Täckström, Andrea Esuli, and Fabrizio Sebastiani. 2014. Hierarchical multi-label conditional random fields for aspect-oriented opinion mining. In *Proceedings of the European Conference on Information Retrieval*. 273–285.

[92] George Markopoulos, George Mikros, Anastasia Iliadi, and Michalis Liontos. 2015. Sentiment analysis of hotel reviews in Greek: A comparison of unigram features. In *Cultural Tourism in a Digital Era*. Springer Proceedings in Business and Economics. Springer, 373–383.

[93] Gustavo Soares Martins, Alcione de Paiva Oliveira, and Alexandra Moreira. 2017. Sentiment analysis applied to hotels evaluation. In *Proceedings of the International Conference on Computational Science and Its Applications*. 710–716.

[94] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal* 5, 4 (2014), 1093–1113.

[95] Fuad Mehraliyev, Irene Cheng Chu Chan, and Andrei Petrovich Kirilenko. 2022. Sentiment analysis in hospitality and tourism: A thematic and methodological review. *International Journal of Contemporary Hospitality Management* 34, 1 (2022), 46–77.

[96] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arxiv:1301.3781* (2013).

[97] Christopher E. Moody. 2016. Mixing Dirichlet topic models and word embeddings to make lda2vec. *arXiv preprint arXiv:1605.02019* (2016).

[98] Putra Fissabil Muhammad, Retno Kusumaningrum, and Adi Wibowo. 2021. Sentiment analysis using Word2vec and long short-term memory (LSTM) for Indonesian hotel reviews. *Procedia Computer Science* 179 (2021), 728–735.

[99] Rizka Putri Nawangsari, Retno Kusumaningrum, and Adi Wibowo. 2019. Word2Vec for Indonesian sentiment analysis towards hotel reviews: An evaluation study. *Procedia Computer Science* 157 (2019), 360–366.

[100] Ambreen Nazir, Yuan Rao, Lianwei Wu, and Ling Sun. 2022. Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. *IEEE Transactions on Affective Computing* 13, 2 (2022), 845–863.

[101] Hoang-Quan Nguyen and Quang-Uy Nguyen. 2018. An ensemble of shallow and deep learning algorithms for Vietnamese sentiment analysis. In *Proceedings of the 2018 5th NAFOSTED Conference on Information and Computer Science (NICS'18)*. IEEE, Los Alamitos, CA, 165–170.

[102] Peter Oram. 2001. *WordNet: An electronic lexical database*. Christiane Fellbaum (Ed.). Cambridge, MA: MIT Press, 1998. Pp. 423. *Applied Psycholinguistics* 22, 1 (2001), 131–134.

[103] John Pavlopoulos and Ion Androutsopoulos. 2014. Aspect term extraction for sentiment analysis: New datasets, new evaluation measures and an improved unsupervised method. In *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM'14)*. 44–52.

[104] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP'14)*. 1532–1543.

[105] Jantima Polpinij, Natthakit Srikanjanapert, and Paphonput Sopon. 2017. Word2Vec approach for sentiment classification relating to hotel reviews. In *Proceedings of the International Conference on Computing and Information Technology*. 308–316.

[106] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16)*. 19–30.

[107] Puteri Prameswari, Isti Surjandari, and Enrico Laoh. 2017. Opinion mining from online reviews in Bali tourist area. In *Proceedings of the 2017 3rd International Conference on Science in Information Technology (ICSITech'17)*. IEEE, Los Alamitos, CA, 226–230.

[108] Reza Amalia Priyantina and Riyanarto Sarno. 2019. Sentiment analysis of hotel reviews using latent Dirichlet allocation, semantic similarity and LSTM. *International Journal of Intelligent Engineering and Systems* 12, 4 (2019), 142–155.

[109] Pradeep Racherla, Daniel J. Connolly, and Natasa Christodoulidou. 2013. What determines consumers' ratings of service providers? An exploratory study of online traveler reviews. *Journal of Hospitality Marketing & Management* 22, 2 (2013), 135–161.

[110] Devakunchari Ramalingam and Valliyammai Chinnaiah. 2018. Fake profile detection techniques in large-scale online social networks: A comprehensive review. *Computers & Electrical Engineering* 65 (2018), 165–177.

[111] Toqir Ahmad Rana, Yu-N. Cheah, and Sukumar Letchmunan. 2016. Topic modeling in sentiment analysis: A systematic review. *Journal of ICT Research and Applications* 10, 1 (2016), 76–93.

[112] Biswarup Ray, Avishek Garain, and Ram Sarkar. 2020. An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. *Applied Soft Computing* 98 (2020), 106935.

[113] Antonio Reyes and Paolo Rosso. 2012. Making objective decisions from subjective data: Detecting irony in customer reviews. *Decision Support Systems* 53, 4 (2012), 754–760.

[114] Ana Reyes-Menendez, Jose Ramon Saura, and Ferrão Filipe. 2019. The importance of behavioral data to identify online fake reviews for tourism businesses: A systematic review. *PeerJ Computer Science* 5 (2019), e219.

[115] Francesco Ricci and René T. A. Wietsma. 2006. Product reviews in travel decision making. In *Information and Communication Technologies in Tourism 2006*. Springer, 296–307.

[116] Sebastian Ruder, Parsa Ghaffari, and John G. Breslin. 2016. INSIGHT-1 at SemEval-2016 Task 5: Deep Learning for Multilingual Aspect-Based Sentiment Analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16)*. 330–336.

[117] Raksmey Sann and Pei-Chun Lai. 2020. Understanding homophily of service failure within the hotel guest cycle: Applying NLP-aspect-based sentiment analysis to the hospitality industry. *International Journal of Hospitality Management* 91 (2020), 102678.

[118] Sergej Schmunk, Wolfram Höpken, Matthias Fuchs, and Maria Lexhagen. 2013. Sentiment analysis: Extracting decision-relevant knowledge from UGC. In *Information and Communication Technologies in Tourism 2014*. Springer, 253–265.

[119] Kim Schouten and Flavius Frasincar. 2015. Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering* 28, 3 (2015), 813–830.

[120] Markus Schuckert, Xianwei Liu, and Rob Law. 2016. Insights into suspicious online ratings: Direct evidence from TripAdvisor. *Asia Pacific Journal of Tourism Research* 21, 3 (2016), 259–272.

[121] Ali Akbar Septiandri and Arie Pratama Sutiono. 2019. Aspect and opinion term extraction for aspect based sentiment analysis of hotel reviews using transfer learning. *arXiv preprint arXiv:1909.11879* (2019).

[122] Glenn Shafer. 1976. *A Mathematical Theory of Evidence*. Vol. 42. Princeton University Press, Princeton, NJ.

[123] Han-Xiao Shi and Xiao-Jun Li. 2011. A sentiment analysis model for hotel reviews based on supervised learning. In *Proceedings of the 2011 International Conference on Machine Learning and Cybernetics*, Vol. 3. IEEE, Los Alamitos, CA, 950–954.

[124] Nishit Shrestha and Fatma Nasoz. 2019. Deep learning sentiment analysis of Amazon.com reviews and ratings. *International Journal on Soft Computing, Artificial Intelligence and Applications* 8, 1 (2019), 1–15.

[125] Qianjun Shuai, Yamei Huang, Libiao Jin, and Long Pang. 2018. Sentiment analysis on Chinese hotel reviews with Doc2Vec and classifiers. In *Proceedings of the 2018 IEEE 3rd Advanced Information Technology, Electronic, and Automation Control Conference (IAEAC'18)*. IEEE, Los Alamitos, CA, 1171–1174.

[126] L. B. Shyamasundar and P. Jhansi Rani. 2020. A multiple-layer machine learning architecture for improved accuracy in sentiment analysis. *Computer Journal* 63, 1 (2020), 395–409.

[127] Jasmeet Singh and Vishal Gupta. 2016. Text stemming: Approaches, applications, and challenges. *ACM Computing Surveys* 49, 3 (2016), Article 45, 46 pages.

[128] Aleš Tamchyna and Kateřina Veselovská. 2016. UFAL at SemEval-2016 Task 5: Recurrent Neural Networks for Sentence Classification. In *Proceedings of the International Workshop on Semantic Evaluation*. 367–371.

[129] Duyu Tang, Bing Qin, and Ting Liu. 2015. Deep learning for sentiment analysis: Successful approaches and future challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 5, 6 (2015), 292–303.

[130] Ivan Titov and Ryan McDonald. 2008. Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th International Conference on World Wide Web*. 111–120.

[131] Thang Tran, Hung Ba, and Van-Nam Huynh. 2019. Measuring hotel review sentiment: An aspect-based sentiment analysis approach. In *Proceedings of the International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making*. 393–405.

[132] Alper Kursat Uysal and Serkan Gunal. 2014. The impact of preprocessing on text classification. *Information Processing & Management* 50, 1 (2014), 104–112.

[133] Raisa Varghese and M. Jayasree. 2013. A survey on sentiment analysis and opinion mining. *International Journal of Research in Engineering and Technology* 2, 11 (2013), 312–317.

[134] Ivar E. Vermeulen and Daphne Seegers. 2009. Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management* 30, 1 (2009), 123–127.

[135] S. M. Vohra and J. B. Teraiya. 2013. A comparative study of sentiment analysis techniques. *Journal of Information, Knowledge and Research in Computer Engineering* 2, 2 (2013), 313–317.

[136] Byron C. Wallace. 2015. Computational irony: A survey and new perspectives. *Artificial Intelligence Review* 43, 4 (2015), 467–483.

[137] Hongning Wang, Yue Lu, and Chengxiang Zhai. 2010. Latent aspect rating analysis on review text data: A rating regression approach. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

[138] Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2017. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*.

[139] Xinrui Wang, Jiuxia Sun, and Haizhen Wen. 2019. Tourism seasonality, online user rating and hotel price: A quantitative approach based on the hedonic price model. *International Journal of Hospitality Management* 79 (2019), 140–147.

[140] Mayur Wankhade, Chandra Sekhara Rao Annavarapu, and Mukul Kirti Verma. 2022. CBVoSD: Context-based vectors over sentiment domain ensemble model for review classification. *Journal of Supercomputing* 78, 5 (2022), 6411–6447.

[141] Mayur Wankhade, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni. 2022. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review* 55, 7 (2022), 5731–5780.

[142] Chuhan Wu, Fangzhao Wu, Sixing Wu, Zhigang Yuan, and Yongfeng Huang. 2018. A hybrid unsupervised method for aspect term and opinion target extraction. *Knowledge-Based Systems* 148 (2018), 66–73.

[143] Guangyu Wu, Derek Greene, Barry Smyth, and Pádraig Cunningham. 2010. Distortion as a validation criterion in the identification of suspicious reviews. In *Proceedings of the 1st Workshop on Social Media Analytics*. 10–13.

[144] Zheng Xiang, Zvi Schwartz, John H. Gerdes Jr., and Muzaffer Uysal. 2015. What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management* 44 (2015), 120–130.

[145] Zheng Xiang, Zvi Schwartz, and Muzaffer Uysal. 2015. What types of hotels make their guests (un)happy? Text analytics of customer experiences in online reviews. In *Information and Communication Technologies in Tourism 2015*. Springer, 33–45.

[146] Karen L. Xie, Zili Zhang, and Ziqiong Zhang. 2014. The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management* 43 (2014), 1–12.

[147] Xun Xu and Yibai Li. 2016. The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *International Journal of Hospitality Management* 55 (2016), 57–69.

[148] Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R. Zaiane. 2017. Current state of text sentiment analysis from opinion mining. *ACM Computing Surveys* 50, 2 (2017), 1–33.

[149] Masahiro Yamamoto and Toshiyuki Sekiya. 2019. m_y at SemEval-2019 Task 9: Exploring BERT for Suggestion Mining. In *Proceedings of the 13th International Workshop on Semantic Evaluation*. 888–892.

[150] Qiang Ye, Rob Law, and Bin Gu. 2009. The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management* 28, 1 (2009), 180–182.

[151] Yichun Yin, Yangqiu Song, and Ming Zhang. 2017. Document-level multi-aspect sentiment classification as machine comprehension. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2044–2054.

[152] Yangyang Yu. 2016. Aspect-based sentiment analysis on hotel reviews. *International Journal for Scientific Research and Development* 4, 6 (2016), 726–729.

[153] Lin Yue, Weitong Chen, Xue Li, Wanli Zuo, and Minghao Yin. 2019. A survey of sentiment analysis in social media. *Knowledge and Information Systems* 60, 2 (2019), 617–663.

[154] Ziqian Zeng, Wenxuan Zhou, Xin Liu, and Yangqiu Song. 2019. A variational approach to weakly supervised document-level multi-aspect sentiment classification. *arXiv preprint arXiv:1904.05055* (2019).

[155] Dongsong Zhang, Lina Zhou, Juan Luo Kehoe, and Isil Yakut Kilic. 2016. What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. *Journal of Management Information Systems* 33, 2 (2016), 456–481.

[156] Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8, 4 (2018), e1253.

[157] Qingxuan Zhang and Chongyang Shi. 2019. An attentive memory network integrated with aspect dependency for document-level multi-aspect sentiment classification. In *Proceedings of the Asian Conference on Machine Learning.* 425–440.

[158] Wanting Zhou, Hanbin Wang, Hongguang Sun, and Tieli Sun. 2019. A method of short text representation based on the feature probability embedded vector. *Sensors* 19, 17 (2019), 3728.