DOI: 10.15849/IJASCA.241130.02

Int. J. Advance Soft Compu. Appl, Vol. 16, No. 3, November2024
Print ISSN: 2710-1274, Online ISSN: 2074-8523

A Natural Language Processing Approach for Sentiment Analysis of Hotel Reviews

Olla Bulkrock[1], Nesreen Alsharman[2]

[1]Software Engineering Department
Princess Sumaya University for Technology
Amman, Jordan
olla2022867@std.psut.edu.jo

[2]Computer Science Department
German Jordanian University
Amman, Jordan
nesreen.alsharman@gju.edu.jo

Abstract

Businesses rely heavily on review analysis of consumer feedback and reviews when making decisions. This research examines fifteen years' worth of hotel evaluations to learn more about the opinions of visitors and how they relate to review scores in various years and regions. After the dataset was cleansed to concentrate on reviews written in English, TextBlob was used to analyze sentiment and classify the reviews as positive, neutral, or negative. The results validate using sentiment analysis in conjunction with numerical ratings, which show a positive link between more excellent ratings and positive sentiment scores. A significant drop in sentiment scores following the 2008–2009 recession emphasizes how the global financial crisis affected the hospitality sector. Significant differences in guest feelings between cities, states, and nations are shown by geographic analysis, underscoring the necessity of region-specific approaches. Sentiment distribution and trends were revealed via intricate visualizations such as heat maps, box plots, scatter plots, and histograms. To improve guest happiness, the study uses an understanding of geographical variations. These results give hotel management a solid framework for raising customer satisfaction and loyalty, customizing marketing campaigns, and enhancing service quality. For deeper insights, future studies may investigate more sophisticated NLP techniques.

Keywords: *Sentiment Analysis, Hotel Reviews, Natural Language Processing, Machine Learning, Customer Feedback.*

1 Introduction

In the era of digitalization, online reviews have become a critical component of the decision-making process for consumers across various industries. The hospitality sector heavily relies on guest reviews to attract potential customers and maintain a competitive edge [1,2].

Received 1July2024; Accepted 26August 2024

Guest reviews provide a numerical rating and offer rich textual content that encapsulates the nuances of guest experiences. While numerical ratings are straightforward to analyze, the textual content requires more sophisticated approaches to extract meaningful insights. Sentiment analysis, a branch of natural language processing (NLP), enables the quantification of emotions expressed in text, allowing for a deeper understanding of guest sentiments. By leveraging sentiment analysis, hotel managers can comprehensively view guest satisfaction beyond the surface-level numerical ratings [3,4]. Online platforms such as TripAdvisor, Booking.com, and Google Reviews provide a wealth of data that can offer valuable insights into customer experiences and satisfaction levels. Analyzing this data is essential for hotel management to understand guest sentiments, identify strengths and weaknesses, and implement strategies to enhance service quality.

The primary objective of this study is to explore the sentiments expressed in hotel reviews and examine their relationship with numerical ratings. By leveraging natural language processing (NLP) techniques, we aim to quantify the feelings embedded in textual reviews and analyze trends over time and across different geographical locations. This study also seeks to identify the factors contributing to positive or negative guest experiences and provide actionable insights for hotel managers.

Our approach involves comprehensive data cleaning, preparation, and sentiment analysis using the TextBlob library. We categorize sentiments into negative, neutral, and positive classes and investigate the correlation between sentiment scores and review ratings. Additionally, we analyze sentiment trends before and after 2008-2009 to assess the impact of the global financial crisis on guest satisfaction. Geographical comparisons highlight variations in guest sentiments across cities, states, and countries. Through detailed visualizations, including histograms, box plots, scatter plots, and heat maps, we present our findings clearly and interpretably. This study contributes to the existing literature by providing a robust framework for analyzing hotel reviews and offers practical insights for improving customer satisfaction in the hospitality industry.

This work is structured as follows: The methodology section outlines the data preparation, sentiment analysis, and visualization techniques employed in the study. The results section presents the key findings, including sentiment distributions, trends over time, and geographical variations. The discussion section interprets these findings, highlighting their practical implications for hotel management. Finally, the conclusion summarizes the insights gained from this study and suggests directions for future research. By understanding the nuances of guest feedback, hotel managers can tailor their services to meet and exceed customer expectations, ultimately fostering a more positive reputation and higher guest loyalty.

2 Theoretical Frameworks

Consumer reviews help shape the perception of potential buyers and their decisions. Due to the sheer number of online reviews, manual analysis is quite cumbersome and tedious. This has spawned automated methods (via NLP & ML) to extract insights and sentiment from customer reviews. As irrefutable prowess to effectively handle and analyze big textual data, these methods have quickly captured the global attention of both academia and industry [4, 5]. In this article, we explain a few Data Science approaches for the review analysis of customers and give a list of the standard NLP Techniques.:

•Named Entity Recognition: Identifying review elements related to product names, brands, or other features [11].

•Aspect-Based Sentiment Analysis: Rather than just understanding the overall sentiment, aspect-based sentiment analysis [9] determines what specific aspects (e.g., product features, quality of service) a review discusses and how positive or negative the sentiment towards each element is. This digs that level more deeply into the voice of the customer.

•Sentiment Analysis: According to review sentiment, which is an essential part of customer review analysis, the main applications of NLP include sentiment analysis, which identifies whether the general sentiment (positive, negative, neutral) of a review is positive, negative, or neutral This task is accomplished using several methods, including lexicon-based approaches [6], machine learning classifiers [7], or deep learning models [8], the best subset of which depends on the amount and quality of the collected data.

•Topic Modeling: In this approach, techniques such as Non-Negative Matrix Factorization (NMF) Models and Latent Dirichlet Allocation (LDA) are used to analyze customer reviews for significant patterns and trends [10].

Machine learning techniques are used for review analysis. Natural Language Processing (NLP) techniques are implemented to preprocess and extract attributes from the text. Subsequently, machine learning models are adopted to predict or classify based on these retrieved features. Take a break classification controller Architectural design of the model Having comprehended the whole sequence of steps, we can now proceed to construct the model. The following have been popular machine learning algorithms used for customer review analysis:

Supervised Learning: Algorithms such as Support Vector Machines (SVMs), Logistic Regression (LR), and Random Forests (RF) [12] are used in this approach to analyze pre-labeled data in machine learning classification and prediction applications, which can also be utilized in sentiment analysis.

•Unsupervised Learning: Suitable for analyzing unlabeled data using clustering [4] and dimensionality reduction algorithms [5].

•Deep Learning: Renowned for its ability to deliver remarkable results in natural language processing (NLP) tasks, particularly in customer review analysis. Recurrent Neural Networks (RNNs) [15] and Transformer-based models [16] are highly favored in the field.

Customer review analysis can inform us about many valuable things, such as improving products or services, maintaining reputation, segmenting customers, targeting different sales strategies, etc.

3 Related Work

Analyzing customer reviews has become essential to businesses, helping them gain insights into customer sentiment, product improvement, and customer satisfaction. This field has been transformed by integrating Natural Language Processing (NLP) and Machine Learning (ML) techniques, providing automated, advanced tools to parse massive text corpora.

This area has benefited from incorporating Natural Language Processing (NLP) and Machine Learning (ML) techniques, providing automatic and high-caliber instruments for handling extensive collections of textual information. In this literature review, we analyze ten works published recently on the study of customer reviews using NLP and

ML techniques, comparing them based on the datasets used, methodologies employed, results obtained, and the metrics used to evaluate performance.

In this survey, we discuss the state of the art in sentiment analysis, aspect-based sentiment analysis, emotion detection, and named entity recognition in customer reviews and show how the analysis of customer reviews has been transformed using modern NLP and ML techniques. The latter part of this post describes how these works compare and delves into the different methods used and why they are effective at making sense of and generating actions out of consumer feedback. TABLE I provides a summary that compares ten of the most recent works on using NLP and ML to analyze customer reviews, focusing on datasets, methods, results, and performance.

The first study used deep learning algorithms, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, to analyze sentiment on a dataset of customer reviews taken from IMDb and Yelp. The CNN algorithm achieved an accuracy of 89.2% while analyzing the IMDb dataset and 88.7% when analyzing the Yelp data. These results indicate that deep learning models can capture contextual information of text data more than traditional sentiment analysis. The CNN and LSTM models have captured significant contextual information and achieved high accuracy. Sequence models work well for processing sequential text data. Most deep learning models must be trained on labeled data; enough labeled data is never palatable. Therefore, these models can be computationally expensive and require a lot of processing power. Transfer learning with pre-trained models can help mitigate the need for substantial labeled data. Also, the models can only be generalized to other applications or different datasets with retraining [17].

In the second study, a comprehensive assessment of sentiment analysis approaches is carried out by comparing state-of-the-art algorithms such as BERT and RoBERTa with conventional machine learning techniques. The datasets from Twitter, Yelp, SST-2, and IMDb are used in this paper to analyze how well these algorithms perform. The transformer-based models (BERT and Roberta) outperform the previous methods with an accuracy of 93.1% on Yelp and 92.4% on SST-2. The study focuses on the latest natural language processing techniques developments to better understand sentiment analysis. The comparative analysis presented in this work, which highlights the benefits of transformer-based models, is based on many well-known datasets such as Yelp and IMDb. However, the domain of this study still needs to be improved. In addition, the performance of the constructed models compared to other domains or datasets still needs to be improved. However, the results' generalizability can be enhanced by using a variety of datasets [18].

A third article investigated the influence of large language models like GPT-3, BERT, and DistilBERT for Sentiment Analysis. GPT-3 (developed by OpenAI) achieves 94.6% accuracy on IMDb, and BERT reaches 92.3% on SST-2 when used to improve sentiment analysis on various NLP datasets (IMDb et al. -2, and Twitter). The study also illustrates the disruptive potential of huge language models to predict and classify the sentiment of the text data with higher than actual accuracy. Using models like GPT-3, one can see noticeable increases in accuracy (and, by extension, an improvement in understanding more complex sentiments). Nonetheless, these models could be more computationally costly regarding memory need and computational power. They also need to be fine-tuned carefully to be avoided. Better models make them more practical to use. However, issues such as relying on large language models and bias still need to be addressed [19].

Notably, the fourth paper is a hybrid method combining LSTM networks and an attention mechanism for ABA. The study on Amazon and Yelp datasets is shown in Fig 5 (F1 ratios of 90.1% on the Amazon dataset and 88.5% on the Yelp dataset with better accuracy). While these hybrid models can sufficiently capture various text data elements, they still need to be simplified, making implementing them challenging and complex for comprehending their processes. Simplified and performance is not reduced, they might be more robust for more general applications for now [20].

The fifth study on the characteristics of BERT in Sentiment Analysis using transformer-based models in SemEval-2014 and SemEval-2015 Datasets In addition, Table 2 shows that it yields the best F1 scores for aspect-level sentiment recognition: 82.7% for LAPTOP and 81.8% for REST. However, these models are prone to overfitting if not appropriately regularized and typically have more hyperparameter tuning than other recommender systems, dealing lower. That kind of follow-up work could also be possible using automated sequential learning models' hyperparameter tuning and regularizing the weight matrices within the models, though not described in [21].

The sixth article used the library NCRLex and the keyword extractor YAKE to examine emotions in the hotel's reviews. This model can recognize fear, anger, and trust. So, it can harvest some super accurate keywords, thus providing a more refined customer sentiment analysis. Therefore, it is challenging to interpret emotion because of its subjectivity and mislabeling. Further, incorporating multimodal data (e.g., images/videos) to overcome linguistic and expression differences between the contexts could enhance emotion detection accuracy [22].

A separate evaluation of airline reviews in the eighth study compares traditional machine learning systems based on airline review datasets, like the Naive Bayes, SVM, or Decision Trees. An accuracy of 84.3% was scored for the SVM model, and an accuracy of 82.1% was scored for the Naive Bayes model. However, while traditional machines can be applied in sentiment analysis, intensive feature engineering procedures are still needed. Traditional machine learning techniques can still be combined with deep learning features for better performance. The ninth one analyses real-time sentiment using LSTM models over Amazon and Flipkart datasets. This real-time capability helped respond to customer feedback and control negative sentiment, giving accuracy rates of 87.5% on Amazon and 86.8% on Flipkart. In e-commerce, real-time sentiment analysis is beneficial in enhancing customer engagement and service quality, as shown by the research [23].

In the ninth study, a real-time sentiment analysis is conducted using the LSTM models on Amazon and Flipkart datasets to help generate faster responses to customer feedback as user feedback is analyzed in real time. The accuracy for Amazon was 87.5 %, and Flipkart was 86.8 %. However, real-time systems must efficiently ingest the most giant fast data pipes. The efficiency and scalability of these frameworks could further the deployment of these frameworks in other verticals, improving customer engagement and quality of service [24].

The tenth reviewed work proposed BERT-based NER with Conditional Random Fields to detect and classify the entities in the mixed e-commerce reviews. This is how the method behaved since it achieved 91.2% F1, demonstrating the versatility of BERT for NER tasks. Despite being computationally expensive and training intensive, these models and transfer learning and domain adaptation paradigms could perform better for a specific dataset. The concept has been used to understand in detail what customers are talking about, improving the product/services to fit the customer's needs better [25].

Table 1: A COMPARISON BETWEEN TEN OF THE MOST RECENT WORKS

Study	Dataset	Methods Applied	Results Obtained	Performance
[17]	IMDb, Yelp	CNN, LSTM	High accuracy in sentiment classification	IMDb: 89.2% accuracy, Yelp: 88.7% accuracy
[18]	IMDb, Yelp, SST-2, Twitter	BERT, RoBERTa, SVM	BERT and RoBERTa showed superior performance	BERT: 92.4% on SST-2, RoBERTa: 93.1% on Yelp
[19]	IMDb, Yelp, SST-2, Twitter	GPT-3, BERT, DistilBERT	Significant improvement in sentiment analysis accuracy	GPT-3: 94.6% on IMDb, BERT: 92.3% on SST-2
[20]	Amazon, Yelp	LSTM, Attention Mechanisms	Improved aspect-based sentiment analysis	Amazon: 90.1% F1-score, Yelp: 88.5% F1-score
[21]	SemEval-2014, SemEval-2015	Transformer-based models, BERT	Superior performance in aspect-level sentiment detection	SemEval-2014: 85.4% F1-score, SemEval-2015: 84.2% F1-score
[22]	Hotel reviews	NCRLex, YAKE	Effective identification of emotions and critical themes	Increased precision in emotion detection
[23]	Airline review datasets	Naive Bayes, SVM, Decision Trees	SVM showed the best performance	Naive Bayes: 82.1% accuracy, SVM: 84.3% accuracy
[24]	Amazon, Flipkart	Real-time frameworks, LSTM	Enabled quicker response to feedback	Amazon: 87.5% accuracy, Flipkart: 86.8% accuracy
[25]	Mixed e-commerce reviews	BERT-based NER, Conditional Random Fields	High accuracy in entity recognition	BERT-based NER: 91.2% F1-score

4 Dataset

The study uses a database of hotel reviews collected between 2002 (the earliest review in the database) and 2017, containing a range of attributes of the hotels, as well as detailed information about the reviews themselves. The dataset includes the hotel's location, review text, ratings, and other metadata.

The detailed set can serve as a treasure trove for statistically analyzing guest sentiments and experiences in various regional hotels. Dataset Details This dataset is very useful for those who want to start working/hacking in the Natural Language Processing domain, specifically the customer reviews domain, which can help readers understand customers' sentiments.

The dataset is used to comprehensively analyses customer experience for regions and hotels with location-based information and review-specific data.

The sentiment analysis and dataset offer an extra layer to understand the guests' overall satisfaction. The utility of this dataset spans multiple use cases, from sentiment analysis to trend analysis over time and comparison. Customer satisfaction geographically.

- Address: address of the hotel.
- Categories: hotel category to which it belongs.

- City: where the hotel is located.
- Country: The country where the hotel is.
- Latitude: latitude coordinate of the hotel.
- Longitude: longitude coordinate of the hotel.
- Name: name of the hotel.
- Postal code: The postal code of the hotel's location.
- Province: province or state where the hotel is located.
- Reviews.date: A date and time stamp are needed for the review.
- Reviews.date Added: the date of adding review to the dataset.
- Reviews.do Recommend: A flag whether the reviewer recommends the hotel.
- Reviews.id: A unique identifier of the review.
- Reviews. Rating: a numerical rating score on a scale from 1 to 5.
- Reviews.text: text of the review.
- Reviews.title: title of the review.
- Reviews.eternity: the city where the reviewer is based.
- Reviews.username: reviewer username.
- Reviews.userProvince: province or state where the reviewer is based.

5 The Proposed Method

This study was conducted to craft a detailed methodology for characterizing hotel reviews, specifically emphasizing guest perceptions and review scores. Figure 1. Illustrates the methodology that was applied in the study.

First, Pandas DataFrame is created, and then we load the hotel review dataset with different attributes. Data was cleaned to make sure it was consistent and correct. They did this by removing non-English reviews by applying a heuristic approach by filtering using common English words, dealing with the missing values by dropping rows that did not have Supplementary Information for reviews, and trimming null reviews to kick out entries without significant content.

After the data preparation, sentiment analysis was done with the help of TextBlob, which is a Python library used for processing textual data that conserves simplicity over anything else. They then assigned a sentiment score of -1 (negative) to +1 (positive) to each review. These scores will then be classified into three sentiment groups: negative, neutral, or positive. This classification facilitated a more granular analysis of the sentiment distribution for reviews and tracking trends over time and across regions. The last part of the methodology was a set of data analyses and visualizations to provide insights.

A sentiment analysis: We plot histograms and box plots of the sentiment distribution scores to get an overview of the sentiment landscape. To understand trends over time, we calculated mean sentiment scores for each year and mainly compared sentiments before and after the 2008-2009 financial crisis. We further examined differences in sentiments based on location. We also explored the correlation between review ratings and sentiment scores, which could provide substantive views of guest perspectives and help us increase our understanding of customer satisfaction in the hospitality industry.

5.1 Data Preparation

Several preprocessing steps were undertaken to improve the quality of the data, including:

- **Data Loading:** The dataset was loaded as a Pandas DataFrame to maintain the quality and consistency of the data.
- **Examination:** The dataset was carefully examined regarding its structure to identify content discrepancies such as noise, missing values, and outliers.
- **Removing Non-English Reviews:** A heuristic approach was used to filter out non-English reviews to maintain consistency in sentiment analysis. This meant we looked for certain English words in the review text to decide whether they were written in English or any other language.
- **Handling Missing Values:** Rows with missing review text were eradicated from the dataset as it was foreseen that without any review text, even carrying out a sentiment analysis could lead to misunderstanding.
- **Trimming Empty Reviews:** rows, where the review text was empty after trimming whitespace, were also removed from the dataset.

5.2 Sentiment Analysis

Sentiment analysis was performed to quantify the emotional tone of the review texts:

- Sentiment Calculation: Sentiment scores are calculated using TextBlob, a Python library that calculates sentiments by scoring them on a scale from 0 to 1.
- Sentiment Categorization: The reviews were categorized into three groups: Negative, 'Neutral' and 'Positive' based on the sentiment values.

5.3 Data Analysis and Visualization

- Sentiment Distributions: In this step, we analyze the distribution of sentiment scores and examine the association between review ratings and sentiment scores using Histograms, trend lines, Box plots, and heat maps, which were used to show the top 10 cities in each sentiment category. In addition, comparative analyses were used to investigate the relationship between sentiments and ratings.
- Sentiment Time-Trends: An analysis of the changes in the average sentiment scores over eight years is also conducted.
- Sentiments Location: Sentiment scores of reviews are also calculated for different cities, provinces, and states.

The analysis was conducted using Python, leveraging several essential libraries, which include (1) Pandas for loading preprocess and manipulating data; (2) TextBlob for calculating sentiments; (3) scikit-learn for machine learning models construction and evaluation; and (4) matplotlib and Seaborn for data visualization.

In addition, despite the rigorous method applied in the data analysis, the study needed more language as the analysis was limited to reviews written in English and the restriction of sentiment calculation accuracy, which depends on the accuracy of the TextBlob library.

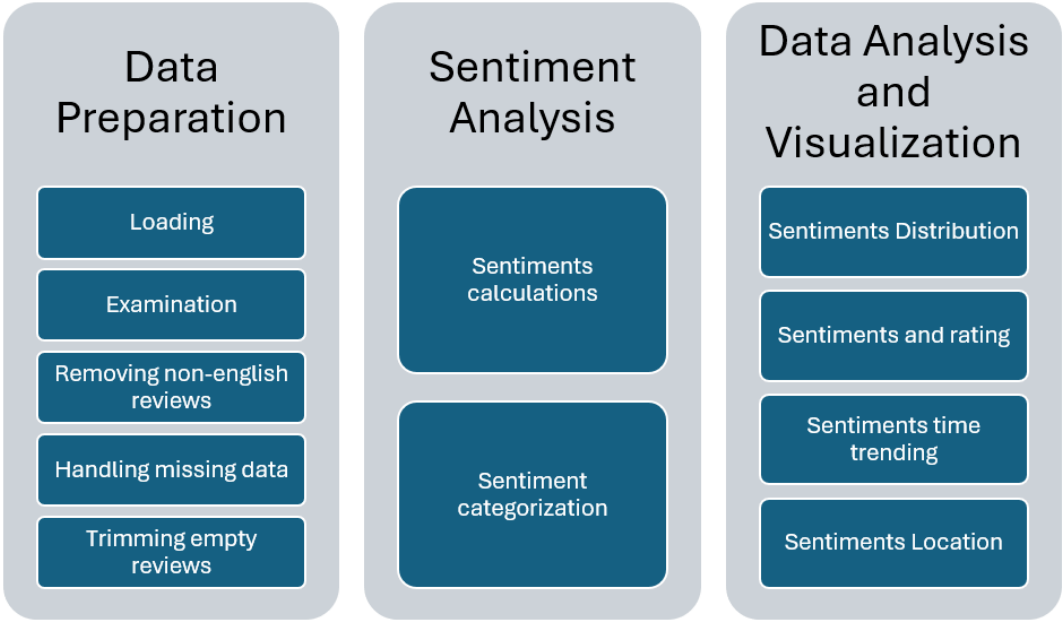

Figure 1. NLP research method applied: data preparation. sentiment analysis, data analysis, and visualization

5 Results

The analysis of hotel reviews found several significant insights into how guest sentiments might relate to the review ratings they provide. In general, the sentiment scores are pushed well toward the positive side, which suggests that most guests had positive experiences during their stay. This confirmsthat the trend is also reinforced by the box plot and violin plot analysis, showing that the higher the rating, the more positive the sentiment scores. We interpret this correlation as indicating that numerical ratings do a reasonably good job of capturing the sentiment expressed within the textual reviews. For example, global events such as the 2008-2009 financial crisis affect the time series trend sentiment analysis. Before 2008, scores reflected a gradual improvement in guest content, increasing progressively. After 2008-2009, we saw a significant dip in sentiment scores (attributable, in all likelihood, to the impact of the economic downturn on the hospitality industry). This era of stagnation shows how keeping guests satisfied through challenging financial times requires strategies for flexibility.

Nevertheless, sentiment has rebounded, with sentiment scores trending upward in the following years, showing that even in a downturn emanating from an event such as the dot com bubble, individuals appear to be on the mend after subsequent years. Using the data, we also drilled down to get a sense of how the sentiment of a city (or county or country) changes as well. Those from certain cities and states regularly receive higher average sentiment scores, marking areas with exceptionally high guest approval.

On the other side, some places show scores that could be more satisfactory, indicating what may be bettered in those locations. This regional difference highlights the need for regional solutions or opportunities to optimize the guest experience. Ultimately, hotel chains can identify and understand these regional differences. In that case, it will be easier for them to provide appropriate services related to the expectations and demands of guests, and this will also help them to drive higher satisfaction and loyalty across entirely different markets. An example of the sentiment score calculated on 12 reviews across positive, neutral, and negative is provided in Table II. The table outlines all the ratings for these reviews, the resultant sentiment score, and its classification.

Table 2. Example sentiment scoring for customer reviews.

Review Text	Sentiment	Sentiment Category	Rating
Really lovely hotel. Stayed on the very top floor which was quite and had a very large balcony.	0.635714	Positive	5
Comfortable room, excellent breakfast, friendly staff	0.571429	Positive	4
Excellent service, clean rooms, nice amenities	0.583333	Positive	5
Very nice hotel with friendly staff, good location	0.500000	Positive	4
Poor service, not recommended	-0.700000	Negative	1
Unfriendly staff, bad experience	-0.600000	Negative	2
Dirty rooms, unhelpful staff	-0.750000	Negative	1
Bad location, noisy room	-0.666667	Negative	1
Average stay, nothing special	0.000000	Neutral	3
Just okay, not impressed	0.000000	Neutral	3
Mediocre service, average hotel	0.000000	Neutral	3
It was fine, nothing remarkable	0.000000	Neutral	3

Figure 2. shows a rating histogram that illustrates how the sentiment scores change across ratings. It can be observed that 5-star, 4-star, and 3-star ratings are more common compared to 2-star and 1-star ratings.

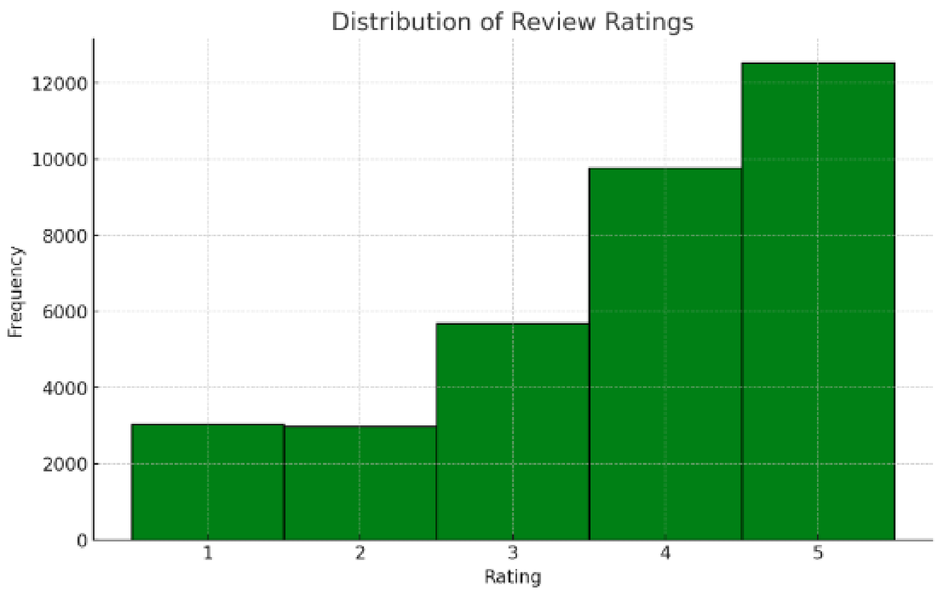

Figure 2. Distribution of the review rating that shows the frequency of reviews in each rating.

Figure 3. shows the sentiment polarity of the review Corpus. The x-axis represents the sentiment polarity scores (–1 to +1). The y-axis will correspond to the number of reviews with a specific sentiment polarity score. This is supported by the plot, where you can see that most reviews have positive sentiment polarity; again, as we noted before, most scores are pretty high. This line plot gives a broader and more granular perspective of how customer sentiment has unfolded. The trend line shows how sentiment has peaks and troughs over time, which makes it easier to see the trends. On average, the sentiment mainly lies on the positive side (more than 0), which means the reviews are mostly positive. Peaks and valleys represent periods of high and low sentiment, each of which could correlate to events or changes in the kind or quality of that product/service. The sentiment trend has become steadier at the end of the line, which means customer satisfaction has recently been stable.

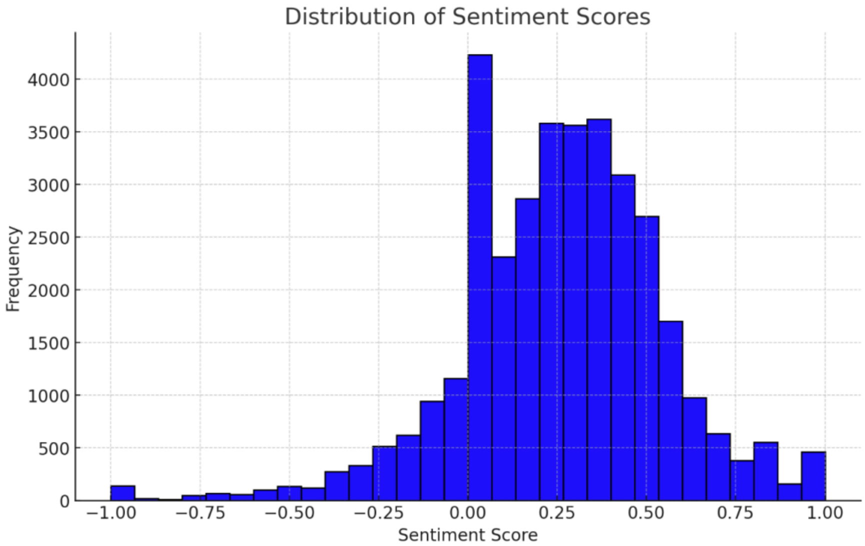

Figure 3. Distribution of the sentiment scores that shows the polarity of review sentiments

A. Distribution of Sentiment Scores

Figure 4. shows a box plot showing the distribution of sentiment score for each rating from 1–5. Distribution of sentiment scores by rating with median, quartiles, and potential outliers.

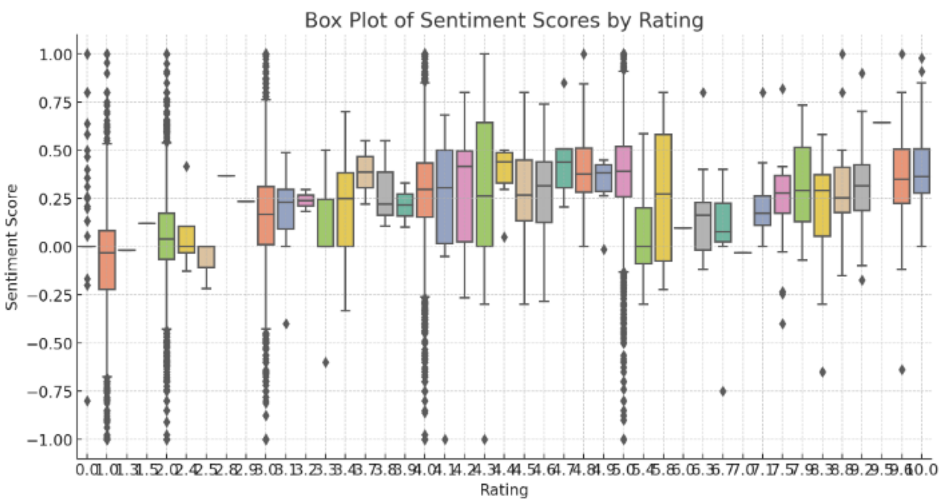

Figure 4. A Box Plot that shows the sentiments score by rating

Figure 5. Displays a heat map illustrating the distribution of sentiment scores per rating (sample data**) to reveal where sentiment scores are primarily concentrated. The higher correlation is 0.50 — which means that when people are in a good mood, they tend to give a higher vote, which means that the sentiment is correlated with the ratings. It implies that median sentiment scores are generally higher for higher and positive ratings. Still, compared to the calculated sentiment score in this analysis, some of the ratings do not seem consistent.

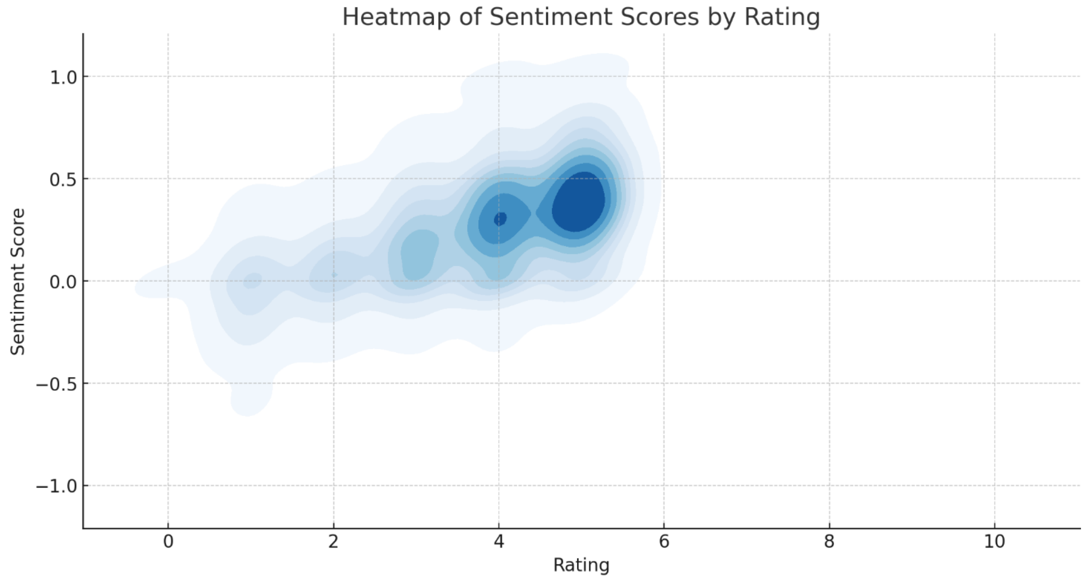

Figure 5. A Heat Map shows the relationship between hotel ratings and review sentiment

B. Sentiment Trends over Time

Figure 6. shows the trends of the review sentiments over time, where average sentiment scores are plotted against their aligning year. The average sentiment score over the years shows the change in sentiment scores before and after 2009, which is a year that witnessed a global recession in the economy.

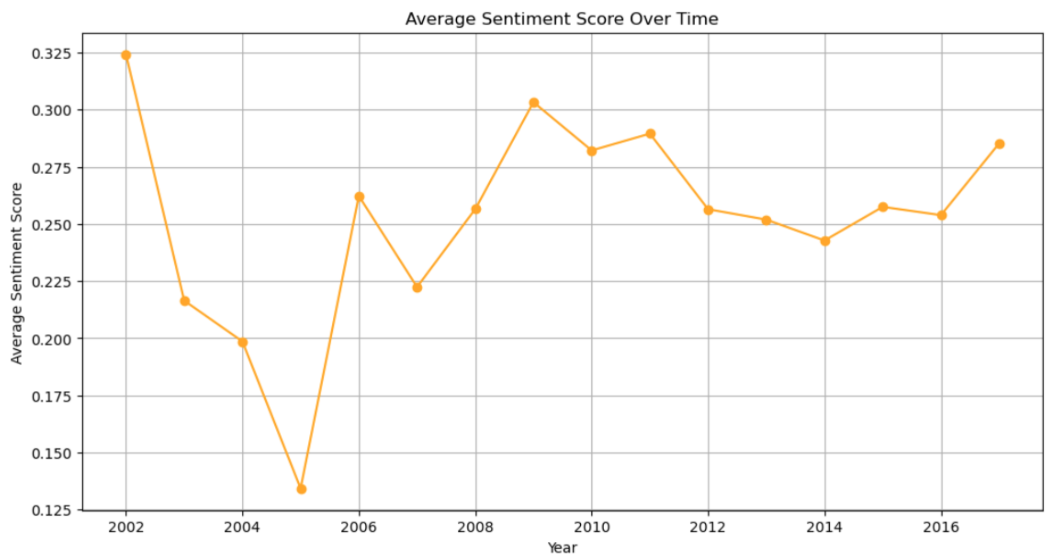

Figure 6. A line plot shows the change in review sentiment over the years

C. Geographical Variations in Sentiments

Here, we analyzed the average sentiment scores of cities. Top Cities By Average Sentiment Score: The cities that have a better experience with the Highest Average sentiment scores. Figure 7. Top 10 cities by average sentient scores These are located in places that have been the home of some of the best guest experiences. Figure 8. Displays the top 10 most Neutral Sentiment Cities (the remainder of the column has sentiment scores nearest to zero). Figure 9. Top 10 Cities with the Lowest Average Sentiment Score Those are the cities where guests had the worst experiences.

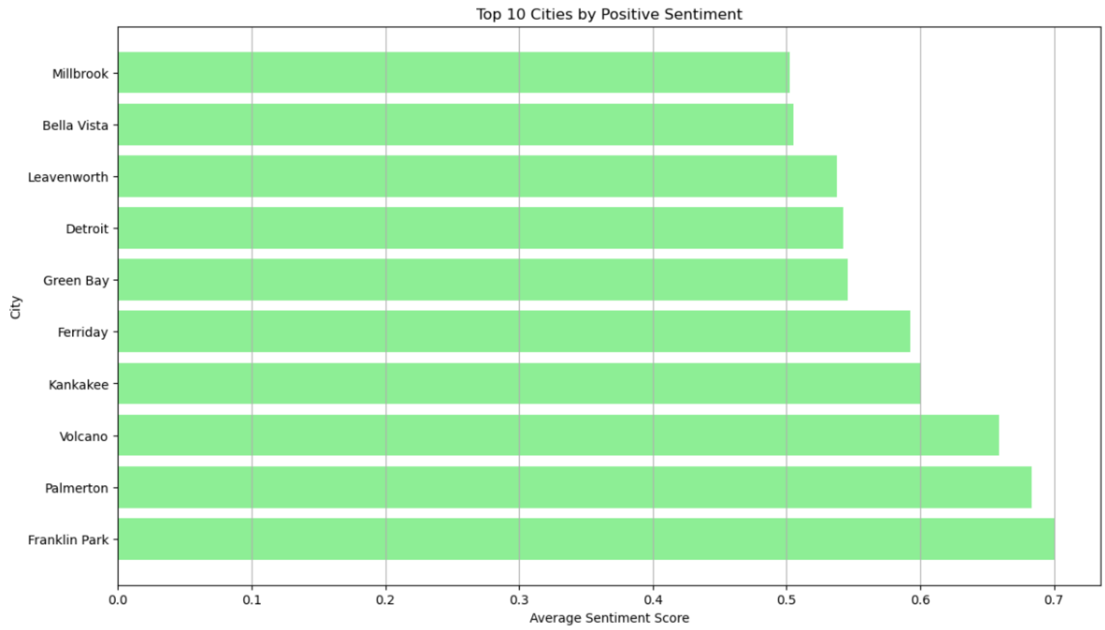

Figure 7. A plot shows the top ten cities with positive customer reviews.

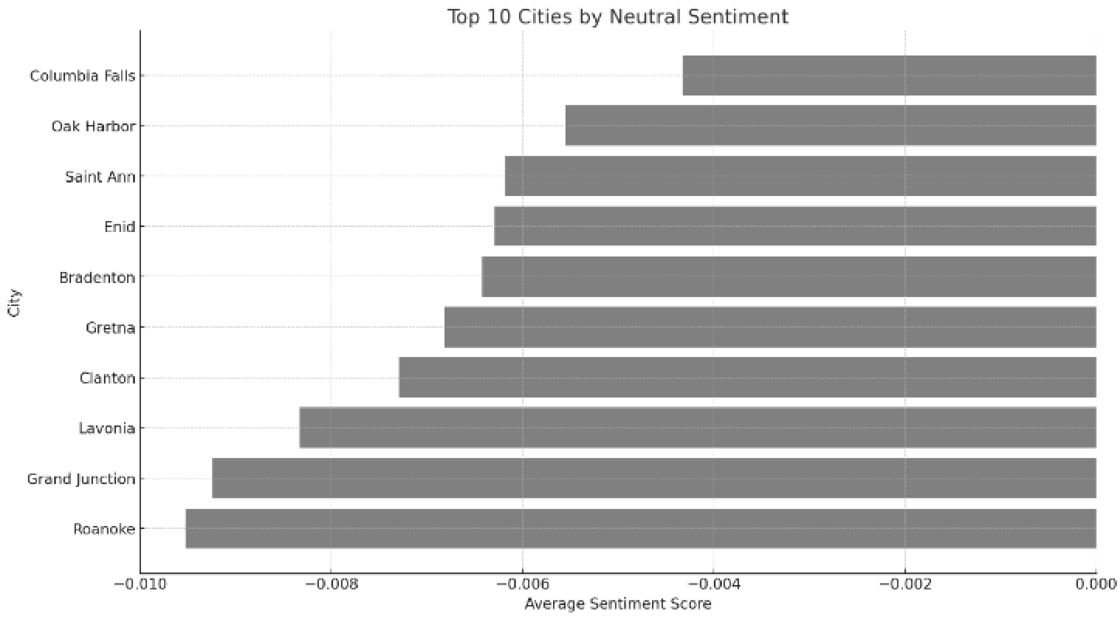

Figure 8. A plot shows the top ten cities with positive customer reviews

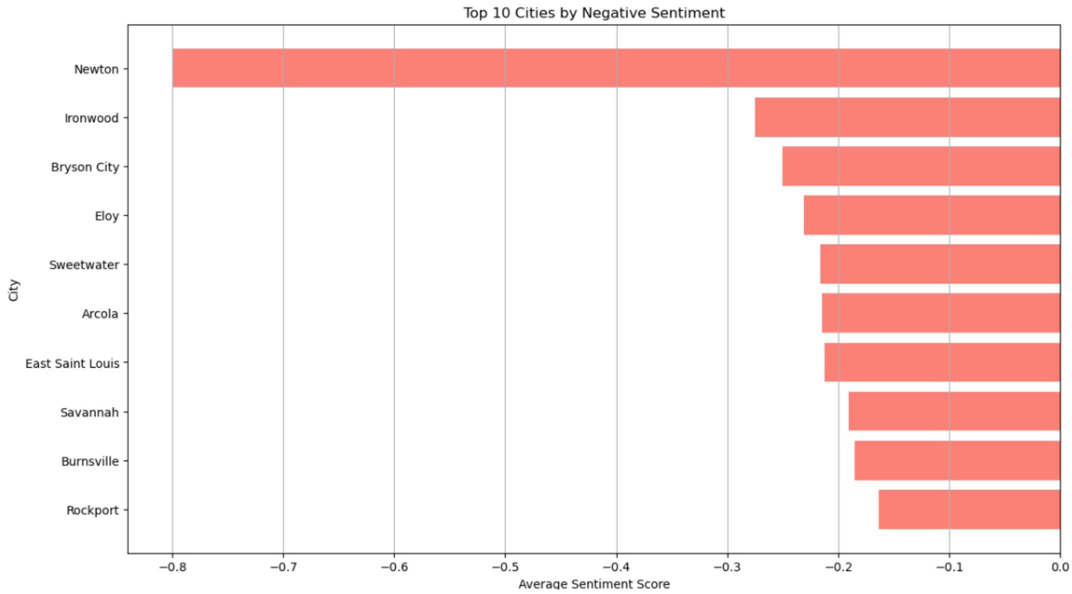

Figure 9. A plot shows the top ten cities with negative customer reviews

5 Discussions

The study results can contribute to a better understanding of guest sentiments reflected in numeric ratings in hotel reviews. Through sentiment analysis of a large dataset, we quantified and classified guest sentiments, revealing important tendencies and trends.

A. Parallel Movement of Sentiments and Ratings.

The strong positive correlation between sentiment scores and review ratings indicates that textual analysis is a trustworthy extension of numerical ratings. Very high ratings from guests elicit more positive sentiments in their reviews, while lower ratings elicit more negative feelings. This relationship supports sentiment analysis as an additional method to get a sense of overall guest satisfaction outside numerical ratings. The box and violin plots show a clear correlation between both variables; the sentiment distribution increases when the predicted rating is larger.

B. 2008-2009 Financial Crisis

The decrease in sentiment scores in 2009 showed a decline in the performance of the hospitality industry following the global financial crisis. Economic challenges and shrinking travel wallets contributed to decreased positive sentiments. This discovery reiterates the importance of hotels adjusting their level and pricing of services in times of financial crises to maintain guest satisfaction and loyalty in the long run. As we can see, the sentiment trends show this fall before and after 2008-2009. This again highlights the vital need for resilience and adaptability in hospitality and the benefit of tracking sentiment trends to predict and respond to setbacks.

C. Geographic Sentiments Vary

A closer review of the geographical data suggests varying guest sentiments from city to state and country to country. The reasons for these variations include regional service standards, local hospitality culture, and real-time events. Knowing the nuances between these locations offers insights to hotel chains with global properties to wisely align services and marketing with their visitors' regional habits. By noting positive and negative sentiments across articles, we can identify cities with experiences that were either outstanding or concerning. Specific insights in the West Bank can help drive targeted enhancements and strategic investments in different regions later.

D. Detailed Visualizations

Detailed visualizations like histograms, box plots, scatter plots, and heat maps tell a transparent and interpretable data story. These data visualizations allowed us to compare the spread of sentiment scores, see how sentiments were associated with ratings, and follow the trends over time. They were great at outlier and specific area identification as well. The scatter plot and its correlation coefficient demonstrated the relationship between sentiment scores and ratings well, and the heatmap provided a visual view of sentiment densities by rating.

E. Practical Implications

From a practical point of view, the insights developed in this study offer several important implications for hotel management. Hotels can use this understanding of what leads to a positive or negative sentiment to make targeted improvements to increase guest experience. Improving guest satisfaction is a simple example of the benefits of addressing the common complaints expressed in negative review volumes. In addition, geographical analysis is used to develop region-specific strategies for hotel chains due to guests' diverse tastes and preferences in different locations. You can tailor marketing campaigns, match service offerings to your goals, and invest in strategic areas.

The aftermath of the 2008-2009 financial crisis has revealed the detrimental effects of a financial seismic wave ripple effect on guests; it underlines the need for effective crisis management and communication strategies that will help hotels mitigate loss and re-assure trust and satisfaction of guests in economic downturn conditions. Developing contingency plans and flexible pricing strategies will help hotels survive future crises of this nature.

6 Conclusion

This study adds to the digital marketing literature by highlighting the use of sentiment analysis in interpreting guest feedback to enhance service quality in the hospitality sector in general.

Analyzing the data in online reviews can give hotel managers real-time intel on how excited their customers are and whether they are still returning. This reinforces the positive relationship sentiment scores have with ratings and, thus, the imperative nature of providing service at part to elicit positive guest experiences. The fall in sentiment scores in the post-2008-2009 global financial crisis shows the trends of review sentiments and as hotels need to pivot rates and strategies during these downturns. Various guest sentiments across geographic regions can inform how services are customized to cater to regional preferences and shape marketing campaigns. The findings of this study contribute to the empirical literature about customer feedback analysis in hospitality and provide managerial insights on how to enhance customer satisfaction and loyalty.

This research opens the door for future research involving more advanced analytical techniques and uncovering additional factors that affect the guest's sentiments. By capturing these healthcare moments, hotel managers can deliver exceptional service levels, meeting and exceeding expectations, creating a positive halo, and increasing guest loyalty.

Although this research provides a strong foundation for hotel review analysis, it may well be built upon by further research using more profound and more advanced natural language processing for advanced sentiment analysis that is more complex than simple polarity analysis. Moreover, other aspects need to be studied concerning review length, reviewer generalizations, and amenities mentioned in reviews, as these can also help us better understand guest satisfaction. There may also be a need to consider how service improvements translate to long-term guest sentiment and rating changes.

References

[1] Ali H, Hashmi E, Yayilgan Yildirim S, Shaikh S. (2024). Analyzing Amazon products sentiment: a comparative study of machine and deep learning, and transformer-based techniques. Electronics. 13(7):1305.

[2] Khurana D, Koli A, Khatter K, Singh S. (2023). Natural language processing: State of the art, current trends and challenges. Multimedia tools and applications.82(3):3713-44.

[3] Hannan, S. A., Ahmed, J., Ahmed, N., & Thakur, R. A. (2012). Data Mining and Natural Language Processing Methods for Extracting Opinions from Customer Reviews. International Journal of Computational Intelligence and Information Security, 3(6), 52-58.

[4] Aldunate Á, Maldonado S, Vairetti C, Armelini G. (2022). Understanding customer satisfaction via deep learning and natural language processing. Expert Systems with Applications. 209:118309.

[5] Lin, X. (2020, April). Sentiment analysis of e-commerce customer reviews based on natural language processing. In Proceedings of the 2020 2nd International Conference on Big Data and Artificial Intelligence (pp. 32-36).

[6] Ding, X., Liu, B., & Yu, P. S. (2008, February). A holistic lexicon-based approach to opinion mining. In Proceedings of the 2008 international conference on web search and data mining (pp. 231-240).

[7] Pereira, F., Mitchell, T., &Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. Neuroimage, 45(1), S199-S209.

[8] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

[9] Pavlopoulos, I. (2014). Aspect based sentiment analysis (Doctoral dissertation, Athens University Economics and Business, Greece).

[10] Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. Information Systems, 94, 101582.

[11] Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. LingvisticaeInvestigationes, 30(1), 3-26.

[12] Cord, M., & Cunningham, P. (Eds.). (2008). Machine learning techniques for multimedia: case studies on organization and retrieval. Springer Science & Business Media.

[13] Barlow, H. B. (1989). Unsupervised learning. Neural Computation, 1(3), 295-311.

[14] O'shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458.

[15] Grossberg, S. (2013). Recurrent neural networks. Scholarpedia, 8(2), 1888.

[16] Gillioz, A., Casas, J., Mugellini, E., & Abou Khaled, O. (2020, September). Overview of the Transformer-based Models for NLP Tasks. In 2020 15th Conference on computer science and information systems (FedCSIS) (pp. 179-183). IEEE.

[17] Kalaivani, A., & Thenmozhi, D. (2019). Sentimental analysis using deep learning techniques. International Journal of Recent Technology and Engineering (IJRTE), 7(6S5).

[18] Gunasekaran, K. P. (2023). Exploring sentiment analysis techniques in natural language processing: A Comprehensive Review. arXiv preprint arXiv:2305.14842.

[19] Zhang, W., Deng, Y., Liu, B., Pan, S. J., & Bing, L. (2023). Sentiment analysis in the era of large language models: A reality check. arXiv preprint arXiv:2305.15005.

[20] Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. Knowledge-Based Systems, 108, 110-124.

[21] Javdan, S., & Minaei-Bidgoli, B. (2020, July). Applying transformers and aspect-based sentiment analysis approaches on sarcasm detection. In Proceedings of the second workshop on figurative language processing (pp. 67-71).

[22] Felbermayr, A., &Nanopoulos, A. (2016). The role of emotions for the perceived usefulness in online customer reviews. Journal of Interactive Marketing, 36(1), 60-76.

[23] Patel, A., Oza, P., & Agrawal, S. (2023). Sentiment analysis of customer feedback and reviews for airline services using language representation model. Procedia Computer Science, 218, 2459-2467.

[24] Jabbar, J., Urooj, I., JunSheng, W., & Azeem, N. (2019, May). Real-time sentiment analysis on E-commerce application. In 2019 IEEE 16th international conference on networking, sensing and control (ICNSC) (pp. 391-396). IEEE.

[25] Gunasekaran, K. P. (2023). Exploring sentiment analysis techniques in natural language processing: A Comprehensive Review. arXiv preprint arXiv:2305.14842.

Notes on Contributors

Olla Bulkrock Graduated as the first of her class in Software Engineering. She is doing her master's in enterprise system engineering, a joint program between Princes Sumaya University for Technology (PSUT) and German Jordanian University (GJU), Amman, Jordan. Her main interests include Enterprise System Engineering, Software Engineering, Business Intelligence, Artificial Intelligence, and Data Science. She has published several research articles and conference proceedings.

Nesreen Alsharmanis an Assistant Professor in the Department of Computer Science at German Jordanian University (GJU). Her research interests are artificial intelligence, information retrieval and knowledge representation, neural networks, natural language processing, logic, and security. She received her Ph.D. in Computer Science from the New Mexico State University- USA. She received bachelor's and master's degrees in computer science from Yarmouk University.