**Deep Survey: Leveraging Neural Networks for Sentiment Analysis, Summarization, & Engagement**

Isaac Caelwaerts

Department of Data Science, University of Wisconsin - Green Bay

DS 785: Capstone

August 1, 2025

**Abstract**

This study investigates the application of deep learning and natural language processing (NLP) to automate analysis of hotel survey responses. Three models were developed to support feedback classification and engagement: a sentiment analysis classifier, a text summarization engine, and an automated reply generator. The sentiment model demonstrated strong performance in identifying extreme feedback—either highly positive or highly negative—but exhibited lower confidence with neutral or mixed responses. To complement this, the text summarization model effectively distilled long-form comments into concise, theme-driven summaries, revealing patterns in guest priorities and expectations. The auto-reply system utilized both outputs to generate contextually appropriate responses, offering hotel staff a scalable solution for real-time engagement. Together, these models illustrate how deep learning can enhance operational efficiency by improving feedback interpretation and streamlining communication. The project provides a foundation for expanding intelligent survey analysis tools in hospitality and other service-focused industries.

*Keywords*: Deep Learning, Natural Language Processing (NLP), Sentiment Analysis, Text Summarization, Automated Engagement, Hotel Survey Responses

**Table of Contents**

## List of Tables

## List of Figures

Chapter 1: Hotels and Surveys

**Introduction**

The Hyatt Hotel Corporation, one of the six major hotel corporations, dominated the American hotel industry. Hotels like it heavily relied on survey responses from their customers to make improvements in their hospitality offerings. Every survey completed by customers presented an opportunity to improve; however, doing so required a painstaking amount of manual review from employees. Countless hours were spent reading through reviews, organizing the data, and creating action plans to resolve issues—often resulting in extended periods of inconsistency.

**Background**

Hospitality professionals working within large hotel chains observed the lengthy and often inefficient process required to translate survey feedback into tangible improvements. Despite the industry's commitment to customer satisfaction, the manual review and implementation cycle frequently led to delays in pragmatic change. Employees tasked with assessing guest reviews had to comb through vast amounts of unstructured text data, organize findings, and formulate response strategies—a process that could take weeks or even months before meaningful adjustments took effect. This inefficiency underscored the need for streamlined, data-driven solutions to enhance decision-making and responsiveness in the hospitality industry.

Customer feedback played a crucial role in shaping the guest experience within the hospitality industry. Hotels like The Hyatt Regency of Green Bay relied on survey responses to identify service gaps, refine operational strategies, and enhance customer satisfaction. However, the process of analyzing feedback remained overwhelmingly manual, requiring employees to sift

through large volumes of unstructured text data to extract actionable insights. This approach was not only time-consuming but also prone to inconsistencies, leading to delayed responses and missed opportunities for meaningful improvements.

Furthermore, traditional feedback systems often failed to capture nuanced sentiment and emerging trends effectively. Without a streamlined approach to categorizing survey data, hotels struggled to prioritize issues and optimize their decision-making processes. As a result, guest concerns frequently went unaddressed, affecting both brand reputation and long-term loyalty. To mitigate these challenges, there emerged a growing need for AI-driven automation in customer feedback analysis. By integrating sentiment analysis, summarization, and fine-tuned chatbot responses, hotels could systematically process survey data, identify patterns, and generate actionable recommendations—enhancing efficiency and ensuring a more proactive hospitality experience.

**Statement of the Problem**

The traditional approach to customer feedback analysis in the hospitality industry was slow, inconsistent, and labor-intensive, preventing hotels from making timely and effective improvements. Employees were required to manually review vast amounts of unstructured survey data, delaying actionable insights and reducing operational efficiency. In addition, conventional methods struggled to capture nuanced sentiment or identify recurring themes, complicating efforts to prioritize concerns or respond effectively. Without an automated system capable of categorizing feedback by sentiment, summarizing common concerns, and generating AI-driven responses, hotels risked losing valuable opportunities to improve guest experiences and strengthen brand loyalty.

**Project Objectives**

This project aimed to modernize customer feedback analysis for the Hyatt Regency of Green Bay by implementing Natural Language Processing (NLP) and deep learning techniques. The dataset consisted of eight years of hotel survey responses, offering a wealth of historical insights into guest sentiment and recurring concerns. The primary objective was to develop a systematic approach that eliminated the inefficiencies of manual review while improving the accuracy and responsiveness of hotel operations.

**Significance of the Project**

To achieve this, the project established an automated pipeline for processing guest feedback. First, an NLP-based sentiment analysis model categorized customer comments according to sentiment, enabling hotel management to quickly identify experiences. This allowed staff to focus on the most critical issues affecting guest satisfaction.

Following sentiment classification, a text summarization algorithm was employed to extract common themes from survey responses, grouping them into recognizable categories such as service quality, cleanliness, booking experience, or staff interactions. This structured organization provided leadership with actionable insights that were immediately applicable to operational improvements.

The final phase of the project involved developing an AI-powered automated response system tailored to customer feedback. Traditional automated responses often lacked empathy, resulting in replies that felt generic, impersonal, or overly rigid. In contrast, this system was designed to learn from staff-edited responses, ensuring that its replies remained natural, thoughtful, and emotionally intelligent. Instead of generating robotic or dismissive messages, the AI was fine-tuned to recognize the nuances of customer sentiment and adapt its responses accordingly—striking a balance between understanding, reassurance, and prompt action. During

the initial implementation, human oversight played a central role in training the system. Over time, the AI refined its approach, ultimately enabling autonomous responses that maintained warmth and sincerity while improving efficiency and guest satisfaction.

The objectives of this project were achieved, and the impact on the hospitality industry, particularly at Hyatt Regency of Green Bay, were substantial. Traditional survey analysis relied on labor-intensive manual reviews, which introduced inconsistency and delayed meaningful improvements to the guest experience. By implementing NLP and deep learning techniques, this project aimed to revolutionize customer feedback workflows, delivering faster, more accurate insights with far less operational strain.

Through sentiment analysis, hotels were able to more efficiently identify critical issues and prioritize concerns that most affected customer satisfaction. Summarization algorithms ensured that recurring themes were properly categorized, enabling leadership to make informed, data-driven decisions based on emerging trends rather than anecdotal evidence. The AI-powered response system, unlike conventional automated replies, generated thoughtful, context-aware messages that aligned with customer sentiment—restoring a sense of personalization to hospitality.

Beyond improving guest satisfaction, the project enhanced business efficiency by reducing manual workload, allowing staff to focus on strategic initiatives and high-value interactions. Over time, as the AI continued to learn from human feedback, the system progressed toward greater automation, ensuring sustained operational efficiency without compromising service quality. Ultimately, this initiative set a new industry standard, showing how AI-driven survey analysis could help hospitality businesses remain agile, responsive, and competitive in an increasingly data-centric world.

The broader impact of this project extended beyond efficiency gains and analytical insights. By streamlining survey processing and automating responses, hotel employees were freed from tedious data tasks—allowing them to focus on what truly mattered: caring for people so they could be their best. This guiding principle of the Hyatt Regency was preserved through the thoughtful integration of technology and human connection, ensuring that service remained warm, personal, and forward-thinking in an AI-enhanced hospitality landscape.

**Assumptions, Limitations, and Delimitations**

The project had clear delimitations to maintain a focused scope centered on survey-based analysis. It did not incorporate customer insights from social media, online reviews, or direct verbal interactions, meaning that broader sentiment data beyond surveys remained outside the project's boundaries. Furthermore, only written text responses were analyzed, excluding feedback that may have come from spoken language or non-verbal cues.

In addition, there may not have been enough data to maximize the AI model's predictive accuracy. To improve performance and generalizability, future iterations of the system might require additional datasets drawn from supplemental sources. Lastly, the AI model was tailored specifically to meet the operational needs of the Hyatt Regency of Green Bay. As a result, it was not designed for universal application across all hospitality brands without further customization. These assumptions, limitations, and delimitations helped frame the project's expectations, ensuring a realistic and effective approach to improving customer feedback analysis within the hotel industry.

**Conclusion**

This project generated operational efficiency in ways previously unattainable. The results of applying NLP and machine learning algorithms offered the potential for immediate action on a

range of issues, some of which may not have been identifiable using manual processes alone. While the Hyatt Regency of Green Bay had already begun exploring data science initiatives, this project provided an opportunity to accelerate that progress by addressing several areas of long standing inefficiency and underdeveloped strategy.

By transforming how the hotel processed and responded to customer feedback, the project created a smarter, faster, and more empathetic service model. Its innovations not only aligned with the goals of modern hospitality but also positioned the Hyatt Regency to lead in adopting AI-driven solutions tailored to real-world operational needs.

Chapter 2: Literature Review on Deep learning and Hotel Surveys

**The Business Problem**

The hospitality industry had long relied on customer satisfaction surveys as a tool for gauging guest experience, informing operational improvements, and guiding strategic decision-making. While it was once sufficient to manually analyze and interpret information from these surveys, new tools had emerged that allowed actionable insights in record time—and even enabled the prediction of future trends. Over the past five years, the adoption of deep learning drastically transformed the landscape of data analysis, pushing the boundaries of what was possible in extracting insights from unstructured sources like surveys. A review of recent academic literature sourced through Google Scholar highlighted how researchers applied deep learning techniques alongside sentiment analysis, text summarization, and AI-generated responses to enhance the interpretation of customer survey data. These approaches aimed to uncover emotional nuance, extract concise insights from unstructured feedback, and generate intelligent responses that mirrored human interaction—particularly within service-driven

industries like hospitality. This research proved vital for understanding the insights deep learning could uncover from survey analysis and identifying the gaps that still remained.

**History of Machine Learning in the Hotel Industry Analysis**

*Application of Machine Learning in the Hotel Industry: A Critical Review* by Alotaibi (2020) explored the gradual integration of AI and machine learning into hotel operations. Alotaibi posed three guiding questions: Where was machine learning being implemented in the hotel industry? What techniques were being used? And which countries were leading in adoption? His findings indicated that, although uptake varied by region, machine learning was most commonly applied to pricing optimization, guest preference modeling, and operational efficiency. However, the review placed limited emphasis on guest feedback systems—particularly surveys. This gap in the literature created an opportunity to examine how deep learning techniques could extract sentiment insights and forecast guest needs from unstructured survey data. Although Alotaibi did not explicitly address the role of deep learning in feedback analysis, this approach offered strong potential for enhancing the operational efficiency he emphasized.

The review identified several popular machine learning use cases, including demand forecasting, dynamic pricing, guest segmentation, and cancellation prediction—all contributing to improved financial and operational outcomes. These applications typically appealed to hotels seeking quick returns and scalable decision-making tools. Large chain hotels, in particular, prioritized demand and pricing forecasts due to the clear business value they provided. Notably, most of the techniques cited in the review predated the widespread adoption of deep learning, which did not gain traction in hospitality research until around 2019. As a result, the industry historically favored traditional algorithms such as decision trees, regression models, and

clustering methods. Geographically, China accounted for approximately 20% of the research and adoption, followed by the United States at 14%. In contrast, Portugal, Spain, and the United Kingdom showed limited engagement, while Middle Eastern countries—particularly those in the GCC—emerged as growing contributors to machine learning research in hospitality.

Alotaibi's conclusion reinforced the idea that, while machine learning held significant promise for the hospitality sector, its current applications remained narrow in scope. The review emphasized the need to address understudied areas such as guest feedback and online review analysis, which could reveal important social influences on customer behavior. By highlighting the growing need for natural language processing and textual feature extraction in review analysis, Alotaibi indirectly affirmed the relevance of applying deep learning to survey data. This project directly addressed that gap by examining how modern AI techniques—particularly sentiment analysis, summarization, and auto-repliers—could enhance hospitality decision-making beyond logistics and pricing. Ultimately, understanding and responding to customer experiences remained central to solving some of the industry's most persistent challenges.

**Sentiment Analysis Research**

Sentiment analysis became a foundational technique within Natural Language Processing (NLP), widely used to extract emotion, opinion, and intent from text. According to Hossen et al. (2021), sentiment analysis was "the most significant to improve a business site," highlighting its importance within hotels. As a result, a growing body of research explored how deep learning–based sentiment analysis of hotel surveys addressed critical business challenges and revealed emerging trends in customer preferences. Studies such as *Hotel Review Analysis for the Prediction of Business Using Deep Learning* (Hossen et al., 2021), *An Improved Model for*

*Sentiment Analysis on Luxury Hotel Review* (Chang et al., 2020), *Performance Comparison of Machine Learning and Deep Learning Models for Sentiment Analysis of Hotel Reviews* (Sanwal et al., 2023), and *A Deep Learning-Based Analysis of Customer Concerns and Satisfaction: Enhancing Sustainable Practices in Luxury Hotels* (Pang et al., 2025) provided insight into how these techniques were deployed to capture guest sentiment and support strategic decision-making in the hospitality industry.

The evolution of machine learning techniques significantly enhanced the accuracy of sentiment analysis in recent years. For example, Hossen et al. (2021) applied a Long Short-Term Memory (LSTM) model to hotel review data and achieved an accuracy rate of 86%, illustrating the effectiveness of deep learning architectures in capturing complex language patterns and emotional nuance in guest feedback. By contrast, Chang et al. (2020) explored sentiment analysis by comparing survey responses with overall review ratings from online book platforms. Their study employed a Random Forest classifier to evaluate the consistency between textual sentiment and numerical ratings, highlighting the model's ability to detect nuanced emotional tone across different feedback formats.

While both traditional machine learning models and deep learning methods had their strengths and weaknesses, research showed that deep learning tended to be more effective at understanding complex language and emotional tone. Models like LSTM and transformer-based systems proved especially adept at capturing how people expressed themselves over time in written feedback. Simpler models like Random Forest or logistic regression often missed those subtle patterns. This shift from easy-to-explain models to more advanced ones reflected a growing focus on accuracy and depth when analyzing large amounts of open-ended survey responses. Since this project aimed to use similar tools for analyzing hospitality surveys, the

strong performance of deep learning models made them a promising option for uncovering meaningful insights and identifying trends in guest satisfaction.

That comparison between traditional and deep learning models was precisely what Sanwal et al. (2023) explored in their performance analysis of sentiment classification tools. Their study directly evaluated classical approaches—such as Logistic Regression, Support Vector Machines, Random Forests, and Decision Trees—against more advanced deep learning models like LSTM and BERT. Using a hotel review dataset, they assessed each model's effectiveness through metrics such as accuracy, precision, recall, and F1-score. The results clearly favored deep learning, with BERT achieving the highest overall performance due to its ability to interpret context and subtle emotional cues. Its bidirectional structure and contextual embeddings enabled it to capture sentiment more accurately than traditional models. LSTM also demonstrated strong results, particularly in handling the sequential flow of customer feedback. These findings strengthened the case for using deep learning in hospitality sentiment analysis, especially when nuanced understanding was key to driving better decisions.

Reinforcing this growing body of evidence, recent research continued to underscore the value of deep learning in understanding guest sentiment. For instance, Pang et al. (2025) applied deep learning to explore how customer sentiment could inform sustainable practices in luxury hotels. Their study analyzed over 29,000 hotel reviews from Henan Province, China, using a combination of topic mining and aspect-based sentiment analysis. This approach enabled them to identify six key areas of customer concern—such as family experiences, service quality, and amenities—and evaluate how each influenced satisfaction at both the overall and aspect-specific levels. Notably, their framework introduced "sentiment quadruples," which captured the category, aspect term, opinion term, and polarity of each review, offering a more nuanced

understanding of guest feedback. By linking sentiment insights to sustainability goals, they demonstrated how deep learning supported not only operational improvements but also long-term strategic planning in hospitality.

Collectively, these studies reinforced the growing shift toward using deep learning as the most robust and accurate method for determining guest sentiment in the hospitality setting. As researchers continued to explore the full potential of these tools, a related focus emerged around *text summarization*—an equally valuable method for distilling lengthy, open-ended survey responses into concise, actionable insights. The following section examined how deep learning was being leveraged not only to detect sentiment but also to automatically summarize guest feedback for faster and more effective decision-making.

**Text Summarization**

Building on the success of sentiment analysis in hospitality research, Bompotas et al. (2020) explored how deep learning could be used not only to classify but also to summarize hotel reviews. Their system categorized feedback by sentiment polarity—positive, negative, or neutral—and generated concise summaries to help businesses quickly interpret large volumes of customer responses. By combining neural network models with opinion mining techniques, the authors demonstrated the value of real-time, sentiment-aware summarization in streamlining decision-making. Importantly, their method also uncovered recurring themes within guest feedback, allowing hotel managers to identify consistent areas of praise or concern. This thematic insight was critical for translating raw sentiment into targeted business actions, reinforcing the role of summarization as both a diagnostic and strategic tool. Their work highlighted a growing shift in hospitality research toward making unstructured feedback both

readable and actionable—further supporting this project's emphasis on applying deep learning to enhance the analysis of open-ended survey responses.

**Automated Replier to Hotel Surveys**

Extending the utility of summarization even further, Ku et al. (2019) examined how deep learning could be harnessed not just to distill feedback, but to guide automated and strategic responses from hotel management. Analyzing over 91,000 guest reviews and 70,000 managerial responses across luxury hotels in London, their study revealed that most hotels lacked a consistent strategy in addressing positive versus negative feedback. To remedy this, the authors developed a deep learning–based prioritization model capable of identifying reviews most deserving of a prompt or customized managerial reply. This approach positioned AI not only as a tool for interpreting sentiment and summarizing themes, but also as a proactive engine for customer engagement and digital service optimization. Their findings demonstrated the expanding role of NLP and visual analytics in operational strategy—an evolution this project aimed to build on by generating dynamic survey responses tailored to guest concerns.

**Conclusion**

Collectively, the reviewed literature revealed that the next wave of actionable insights in the hotel industry would be driven by building upon these deep learning advancements. As researchers moved beyond sentiment detection toward summarization and AI-generated responses, the integration of Natural Language Processing into hospitality operations became increasingly strategic. These technologies not only streamlined the interpretation of guest feedback but also enabled more dynamic, data-informed engagement with customers. This evolving toolkit formed the foundation for this project's objectives: to automate the interpretation

of open-ended survey responses and generate tailored replies that elevated the guest experience and informed business decisions.

Chapter 3: Data Science Application Deep Survey

**Data Science Application**

The objective of this project was to identify and implement the three most effective models for extracting insights from hotel survey feedback: (1) predicting guest sentiment, (2) summarizing guest experiences, and (3) generating personalized auto-replies. Given the nuanced, short-form nature of the feedback collected, this challenge required robust Natural Language Processing (NLP) tools capable of interpreting emotional tone, intent, and contextual subtleties, particularly across varied demographics and writing styles. Although classical machine learning approaches—such as logistic regression, quadratic discriminant analysis (QDA), and decision trees—were considered, transformer-based large language models (LLMs) were ultimately selected. These models supported TF IDF, Zero Shot, few-shot and instruction-tuned classification, abstractive summarization, and contextual reply generation aligned with guest sentiment.

The models explored included Unsloth/Mistral-7B-Instruct-bnb-4bit, which was selected for its lightweight inference performance and flexible prompt design. Additionally, Unsloth/Llama-3-8B-Instruct-bnb-4bit was utilized for its enhanced semantic reasoning capabilities and strong instruction-following performance across both classification and summarization tasks. Each model was evaluated using zero-shot and few-shot prompting techniques to assess its accuracy in sentiment prediction. Gemini 1.5 Flash was also considered for its high-quality generative outputs and refined tone-matching during reply generation; however, it was excluded from final testing due to time constraints, financial limitations, and

free-tier quota restrictions. Text summarization and auto-reply generation were ultimately conducted using Unsloth/Mistral-7B-Instruct-bnb-4bit. Custom prompt functions were optimized to align responses with predicted sentiment classes, ensuring that replies reflected appropriate levels of empathy, gratitude, or resolution, while summaries effectively distilled key themes from each survey in a concise, business-actionable format.

**Data Overview**

Data for this project were obtained directly from the Hyatt Regency of Green Bay's HySat platform, a proprietary system designed to collect and manage guest satisfaction surveys. These surveys spanned a comprehensive time frame, from 2018 through May 2025, and provided a rich and representative sample of guest experiences influenced by seasonal variations, operational shifts, and demographic diversity. To construct a usable dataset, multiple survey exports were compiled and standardized, with particular emphasis on open-text fields such as "Additional Feedback on Overall Stay" and general comment sections. Incomplete and duplicate responses were systematically removed to maintain data integrity.

Textual feedback was segmented and labeled using two distinct approaches: a manually annotated sample of 100 surveys and a larger corpus of over 6,600 responses in which sentiment labels were inferred based on the Likelihood to Return (LTR) score. The LTR metric, rated on a scale from 0 to 10, was mapped to predefined sentiment categories using thresholds derived through exploratory data analysis. This mapping facilitated the identification of patterns between guest sentiment and retention behaviors, distinguishing between minor inconveniences and significant service disruptions. The enriched labeling strategy provided deeper insight into the emotional and operational factors that influenced guest loyalty and revealed key areas for

improvement. The resulting cleaned and structured dataset formed the basis for NLP-driven tasks including sentiment classification, experience summarization, and auto-reply generation.

**Data Cleaning and Processing**

Data for this project was obtained directly from the Hyatt Regency of Green Bay's HySat platform, a proprietary system designed to collect and manage guest satisfaction surveys. The surveys spanned a comprehensive period from 2018 through May 2025 and provided a rich, representative sample of guest experiences influenced by seasonal variation, operational transitions, and demographic diversity. To construct a usable dataset, multiple survey exports were merged into a single DataFrame using Pandas and were subsequently standardized, with particular attention given to open-text fields such as "Additional Feedback on Overall Stay" and general comment sections. Responses containing missing textual feedback were removed using the dropna() method. Although duplicate entries were present, no deduplication was performed, as repeated feedback entries were considered potentially informative for downstream sentiment analysis and behavioral modeling.

Text normalization procedures targeted noisy formatting and encoding inconsistencies. Common HTML artifacts, line breaks, and non-ASCII characters introduced during survey input and export were removed. Lowercasing, whitespace stripping, and formal tokenization were not employed, as early experimentation indicated negligible performance improvement when applied to transformer-based LLMs used in this study.

Sentiment labeling followed two strategies. In the manually annotated sample of 100 responses, labels were assigned via qualitative labeling and grouped into five fixed categories: Very Positive, Positive, Neutral, Negative, and Very Negative—reflecting both emotional tone and perceived guest loyalty. For the full corpus of 6,600+ entries, sentiment labels were inferred

using the Likelihood to Return (LTR) score. The LTR metric, rated on a scale from 0 to 10, was mapped to sentiment categories using thresholds derived through histogram analysis and business heuristics. The resulting mapping was as follows: "Very Positive" (score = 10), "Positive but Critical" (scores 8–9), "Mixed or Neutral" (score = 7), "Slightly Negative" (scores 5–6), "Negative" (scores 3–4), and "Very Negative" (scores 0–2). This approach facilitated correlation modeling between feedback tone and return intention, helping to distinguish minor inconveniences from impactful service disruptions.

To support NLP model deployment, only relevant columns—specifically "Additional Feedback on Overall Stay," the LTR score, and the derived sentiment label—were retained. Personally identifiable information such as names, reservation IDs, and contact details was excluded to ensure privacy compliance.  A post-labeling class distribution analysis revealed moderate imbalance in the larger dataset, with "Very Positive" responses forming the majority class. In contrast, the manually labeled test sample displayed an inverse distribution, with negative sentiment responses dominating the dataset. Although no resampling techniques were applied to address class imbalance, early trial results indicated that classification performance remained relatively stable. Preliminary comparisons between labeling strategies—one derived from Likelihood to Return (LTR) scores and another based on manual emotional annotation—revealed notable differences in prediction consistency and category sensitivity. These differences were not explored in detail within this chapter, but they informed later analysis and shaped the rationale for the selected modeling approaches. This observation also underscored the importance of aligning sentiment labeling methods with the interpretive goals of the feedback system, whether operational, emotional, or predictive in nature.

**Exploratory Data Analysis**

Exploratory data analysis (EDA) was conducted to uncover patterns, trends, and potential modeling challenges prior to training the selected NLP models. This process focused on both quantitative response metrics and qualitative textual patterns derived from the HySat survey data.

The distribution of Likelihood to Return (LTR) scores was examined using a histogram, revealing a heavily right-skewed trend within the large dataset. Most respondents rated their likelihood to return between 8 and 10, with scores of 10 comprising the single most frequent response. This pattern aligned with expectations, given that hospitality survey responses tend to lean positive due to loyalty or satisfaction bias.

**Figure 1** *Distribution of Likelihood to Return (LTR) Scores*
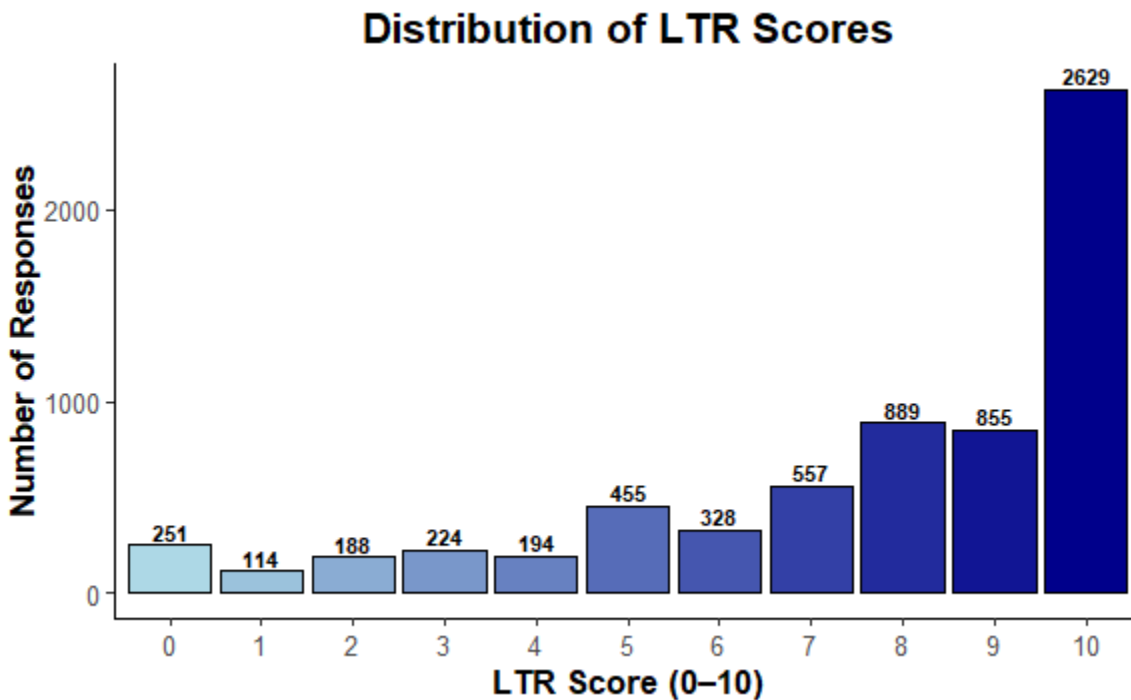


Sentiment label frequencies were analyzed for both the manually labeled test dataset and the LTR-inferred corpus. In the smaller manual sample, negative sentiments dominated, offering a more critical lens for evaluating classifier performance. In contrast, the larger LTR-based

dataset showed a majority of responses classified as "Very Positive" or "Positive but Critical," highlighting natural class imbalance resulting from numerical rating skew.

**Figure 2** *Sentiment Category Frequency — LTR-Inferred Dataset*
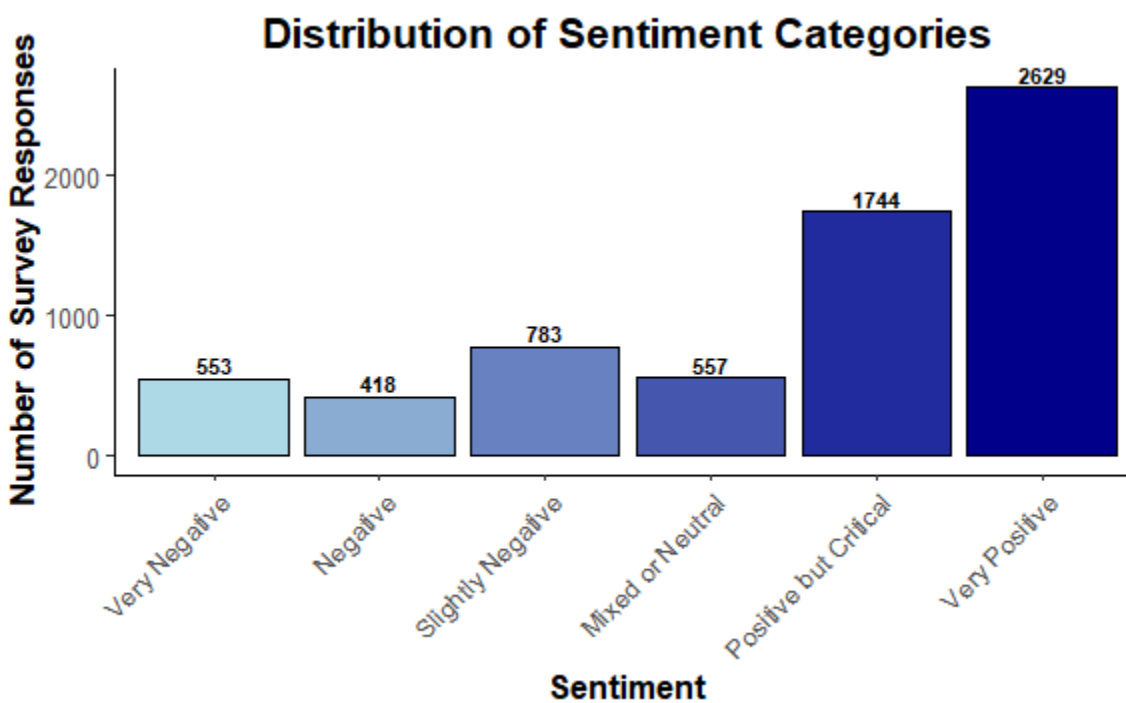
**Figure 3** *Sentiment Category Frequency — Manually Labeled Dataset*
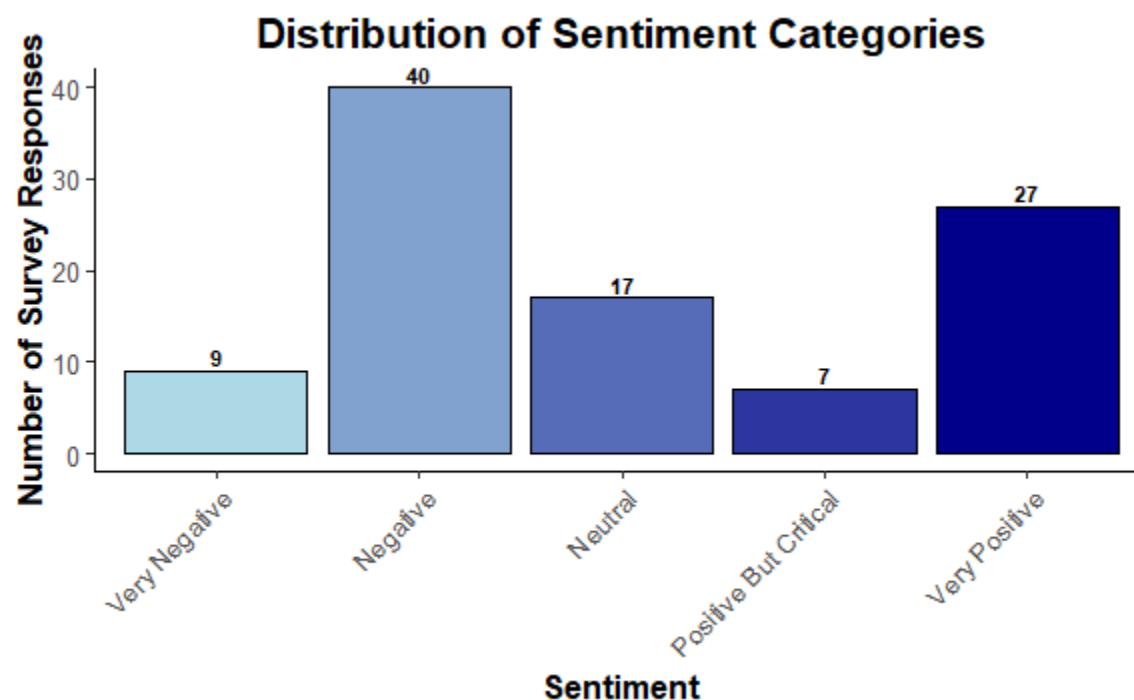


A keyword frequency analysis was performed to identify common terms associated with guest satisfaction and dissatisfaction. Positive feedback frequently included terms such as "clean," "friendly," "convenient," and "quiet," while negative reviews centered around words like "waited," "dirty," "unpleasant," and "never again." These linguistic insights guided prompt engineering and tone-matching logic for downstream summarization and reply generation.

**Figure 4** *Top 20 Most Common Words in Feedback (Stop Words Removed)*

**Top 20 Most Common Words (Stop Words Remove**

| Word | Frequency |
|---|---|
| hotel | 174 |
| staff | 1403 |
| stay | 1214 |
| breakfast | 1088 |
| parking | 1084 |
| desk | 826 |
| check | 788 |
| front | 730 |
| nice | 691 |
| hyatt | 666 |
| food | 662 |
| time | 648 |
| service | 624 |
| night | 582 |
| clean | 445 |
| helpful | 420 |
| friendly | 418 |
| told | 411 |
| bathroom | 408 |
| people | 390 |
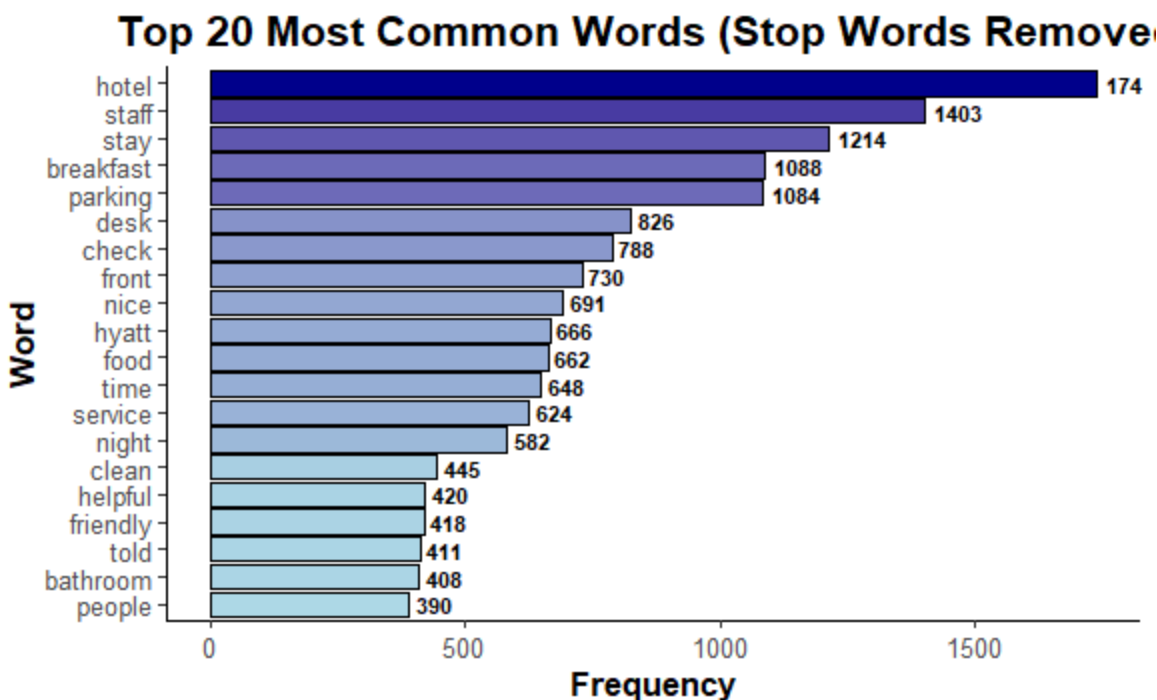
Demographic or seasonal segmentation was not included in this phase of analysis, as the primary modeling objective was to evaluate language-based sentiment patterns using natural language processing techniques. Since the selected NLP models operated solely on textual input, metadata such as guest age, stay type, or visit date was not required to generate sentiment predictions or summaries.

However, this exclusion does not imply that such metadata lacks analytical value. Prior studies suggest that demographic and seasonal factors may influence sentiment expression, linguistic tone, or feedback length. Integrating metadata-aware models may offer richer subgroup-specific insights or expose hidden bias in classification results. Future iterations may incorporate hybrid approaches that combine structured and unstructured data to further refine predictive accuracy and business relevance.

**Model development, Deployment and Evaluation**

The modeling pipeline focused on classifying guest sentiment from survey responses using a combination of traditional machine learning and large language model (LLM) techniques. Two distinct approaches were implemented across both the large LTR-inferred corpus and the smaller manually labeled dataset: a TF-IDF-based logistic regression classifier and prompt-based zero-shot/few-shot classification using instruction-tuned LLMs. Both methods were applied to two datasets — one large corpus inferred from Likelihood to Return (LTR) scores, and one smaller manually annotated test set. For traditional modeling, logistic regression was trained on TF-IDF feature vectors extracted from the preprocessed text. Text cleaning included lowercasing, tokenization, and stopword removal, with hyperparameter tuning focused on optimizing regularization strength and n-gram ranges. Training and validation splits followed an 80/20 structure across all experiments.

In parallel, multiple configurations of prompt-based LLMs were deployed using quantized models such as LLaMA-3 8B and Mistral-7B via the unsloth framework. Both zero-shot and few-shot prompting strategies were tested. Few-shot prompts were dynamically generated from the training data, and later refined with manually selected examples to improve classification performance on edge cases — especially for nuanced categories like "Positive but Critical" and "Mixed." Prompt engineering emphasized strict class boundaries and caution when interpreting polite but critical language. Hyperparameter tuning for these models included adjustments to batch size, token length, and instruction strength through wording and sentiment definitions.

Models were deployed and evaluated within a cloud-based Google Colab environment, allowing for flexible experimentation across both traditional and large language model architectures. Predictions were generated locally using Python-based pipelines, enabling direct

control over input handling, prompt design, and classification logic. These pipelines were intentionally designed to be modular and scalable, supporting future deployment to cloud-based systems such as RESTful APIs or real-time chatbot applications, where guest feedback could be analyzed and responded to automatically. Throughout development, invalid model predictions were logged and excluded, contributing to iterative improvements in prompt engineering. Evaluation metrics included precision, recall, F1-score, and confusion matrix analysis.

**Figure 5**: *Confusion Matrix — TF-IDF Model on LTR-Inferred Dataset*
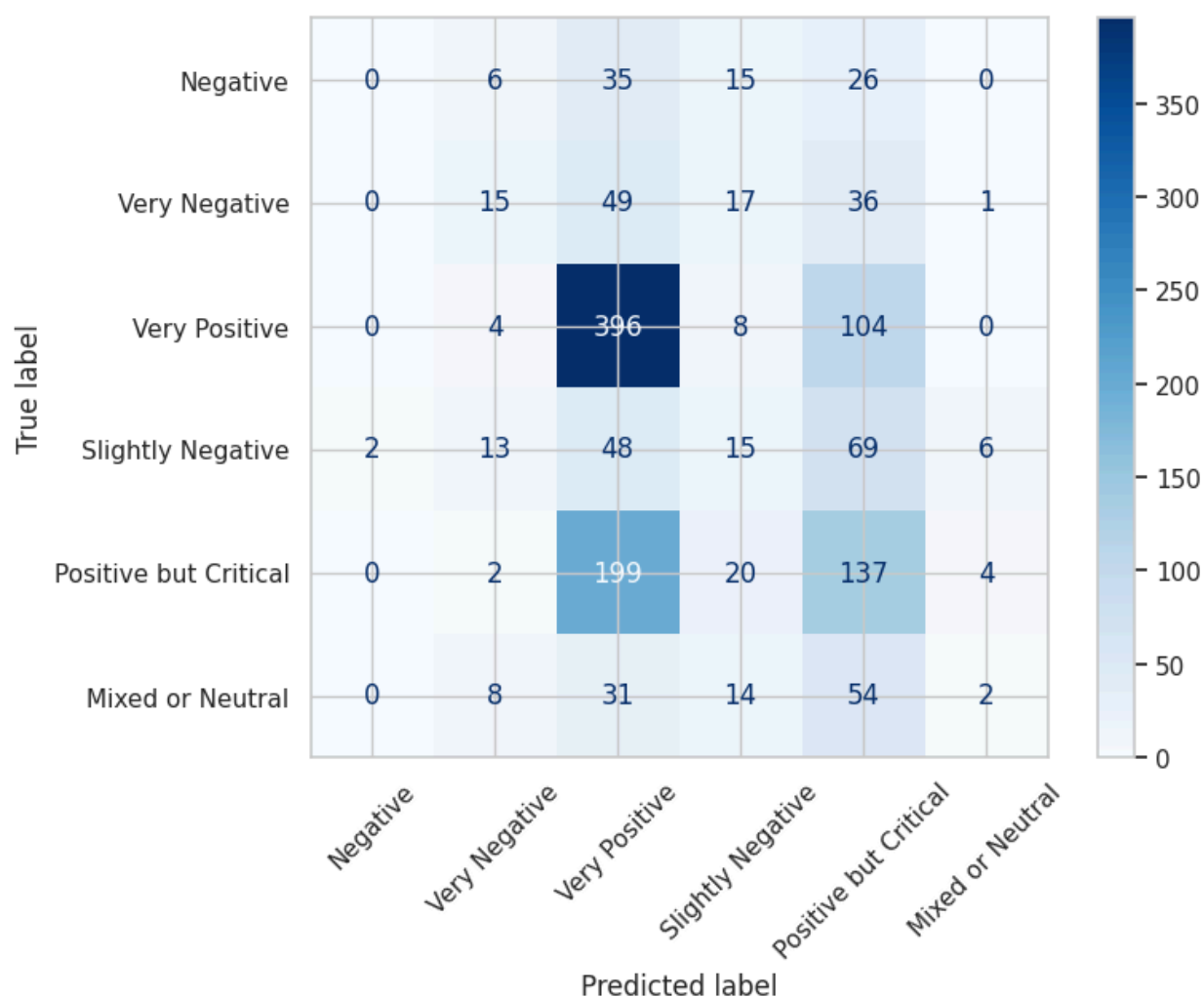
A baseline TF-IDF logistic regression model trained on the large LTR-inferred corpus achieved a weighted average F1-score of **0.37**, with strongest performance observed in polarized sentiment classes such as "Very Positive" and "Very Negative." As illustrated in *Figure 4*, the confusion matrix reveals high precision for dominant categories, while mid-range sentiments like "Mixed or Neutral" and "Slightly Negative" were frequently misclassified. These results reflect the model's sensitivity to class imbalance and its reliance on surface-level lexical cues, which may limit its ability to capture nuanced emotional tone.

**Figure 6**: *Confusion Matrix — Zero-Shot LLM on Manual Dataset*
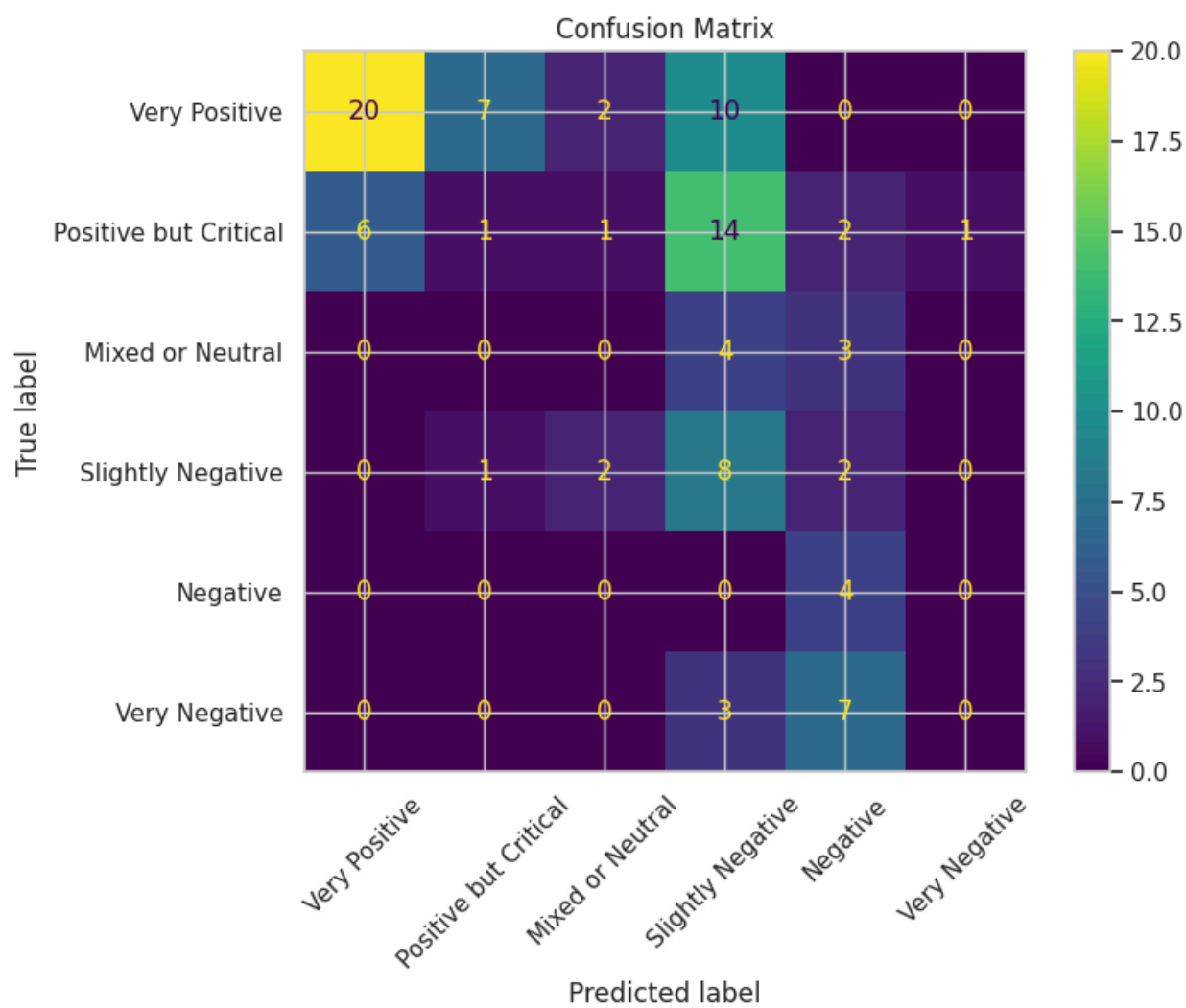
In contrast, zero-shot LLMs tested on the LTR-inferred dataset yielded an average F1-score of approximately 0.32, though they exhibited a tendency to overpredict moderate positivity in ambiguous reviews.

Incorporating few-shot prompts with manually selected edge cases and stricter instructional logic led to measurable improvements in precision for mid-range categories such as "Neutral" and "Positive but Critical." The final LLM handled different types of guest feedback evenly — whether positive, negative, or mixed — without getting stuck or making overly confident guesses. It also worked well with the parts of the system that summarize reviews and write thoughtful replies that match the mood of the guest.

In addition to classification, the project included a suite of text summarization models and tone-aware auto-reply generators. For summarization, three approaches were tested: a Hugging Face pre-trained summarizer, a zero-shot prompting model, and a few-shot prompting model that leveraged curated examples to produce more context-sensitive outputs. Each approach was evaluated for its ability to generate concise, emotionally aligned summaries of guest feedback while preserving core themes and hospitality-specific language. To support both summarization and business intelligence efforts, a custom theme extraction function was developed to identify recurring issues within each LTR-derived sentiment category. These thematic clusters provided operational insight and helped focus hotel management's attention on areas most frequently associated with dissatisfaction or praise. For example, "Mixed or Neutral" responses surfaced themes such as *cold/discomfort/poor quality*, *deterioration*, and *unreasonable fees or poor service*. "Negative" and "Very Negative" sentiments highlighted frustrations like *odor*, *unsatisfactory rooms*, *dirty conditions*, and *poor service quality*, while "Slightly Negative" feedback consistently raised issues with *digital key frustration* and *water-heating inconvenience*.

In contrast, "Very Positive" reviews emphasized strengths such as *cleanliness*, *food variety*, and *pool maintenance*.

Auto-reply generation was implemented via a separate prompting pipeline, designed to generate polite, context-aware responses mapped to each sentiment category. This system used a predefined sentiment-to-tone schema to ensure appropriate levels of empathy, assurance, and gratitude. Replies to negative feedback emphasized accountability and support, while responses to positive feedback reinforced loyalty and appreciation. Although sentiment classification results were not directly used to trigger summarization or reply generation, the consistency in tone and theme across components contributed to a unified guest engagement strategy.

**Conclusion**

Spanning all stages, the data science lifecycle, from data integration and cleaning to model development, deployment, and evaluation, laid the foundation for extracting sentiment, identifying thematic patterns, and generating context-aware responses from hotel guest feedback. These methodological components went beyond producing performance metrics; they enabled operational insight, supported scalable engagement strategies, and facilitated automated communication tools aligned with hospitality service goals. With the pipeline implemented, the results are presented in the following chapter.

Chapter 4: Deep Survey Analysis

**Introduction**

The hospitality industry had long struggled with extracting meaningful insights from guest survey feedback due to the time-intensive nature of manually reviewing open-ended comments. This process often resulted in delayed recognition of operational issues, missed opportunities for service improvement, and underutilization of valuable customer perspectives.

Survey responses—rich in nuance and often emotionally charged—were typically relegated to manual review by hotel staff, making it difficult to scale feedback analysis or respond to recurring themes consistently. To address this challenge, the capstone project applied data science methodologies to automate and enhance the interpretation of guest feedback. The overarching business objective centered on transforming qualitative survey content into actionable intelligence to improve both guest satisfaction and operational efficiency. This objective was addressed through the development of three distinct Natural Language Processing (NLP) models, as outlined in Chapter 3.

The first was a sentiment prediction model designed to evaluate the emotional tone and perceived satisfaction conveyed in guest comments. The second model focused on summarizing extended feedback into concise, digestible formats tailored for managerial review, while also identifying key themes across multiple survey responses. This dual-purpose summarization approach not only reduced cognitive load for hotel staff but also enabled pattern recognition in recurring guest concerns or praise. The third component was an automated reply generator, which produced personalized and sentiment-aware responses to enhance guest engagement and streamline communication workflows. Together, these models formed a scalable pipeline for hotel survey analysis—one capable of converting raw feedback into operational insights and responsive messaging. The following sections evaluate the performance and practical utility of each model in resolving the stated business challenge. In particular, these analyses examine classification accuracy, information compression, thematic extraction, and alignment with customer service goals, ultimately leading to the selection of the most effective configuration for deployment.

**Model 1: Sentiment Analysis**

To assess the emotional tone of guest feedback, a sentiment prediction model was developed using a few-shot prompting strategy powered by the Mistral 7B architecture. This model was designed to classify hotel survey comments into five sentiment categories, ranging from "Very Negative" to "Very Positive." As shown in Figure 5, the model achieved an overall accuracy of 65%, demonstrating its ability to generalize across a diverse range of guest responses. The classification report revealed high precision in detecting "Very Positive" comments (1.00), along with strong recall for both "Very Negative" (0.78) and "Positive but Critical" categories (0.86). These metrics indicated the model's effectiveness in extracting emotionally charged and actionable feedback. While performance was more moderate in the "Neutral" and mixed-sentiment classes, the weighted average F1 score of 0.67 suggested balanced classification across categories. The use of explicit label definitions and decision criteria likely contributed to the model's interpretive consistency. Figure 5 presents both the classification metrics and confusion matrix, offering a detailed visualization of the model's performance. These results support its utility as a robust tool for automating sentiment analysis in hospitality survey workflows.

**Figure 7 Few-Shot Confusion Matrix**

Few-Shot Confusion Matrix
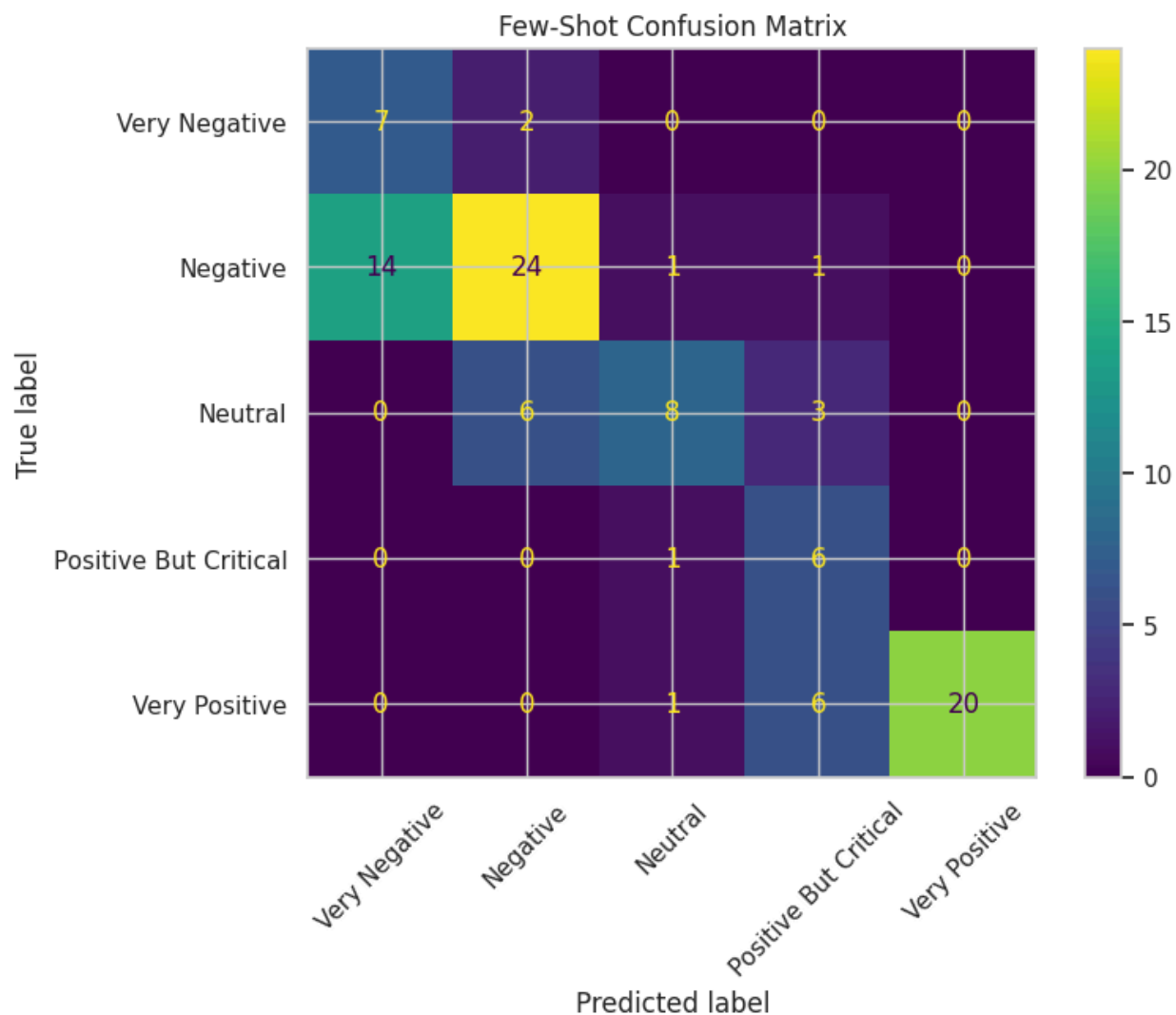
The results presented in Figure 5 reinforced the model's suitability for operational deployment, particularly in its ability to identify satisfaction extremes that warranted immediate attention. For example, the high recall associated with "Very Negative" feedback enabled hotel staff to prioritize and respond to severe guest complaints in a timely manner, while the model's precision in classifying "Very Positive" responses facilitated trend analysis related to service excellence. Although accuracy in detecting "Neutral" and mixed-sentiment comments remained moderate, the classifier consistently added structure to otherwise unstructured survey data. Its application streamlined the feedback triage process and provided managers with

sentiment-driven insights that informed staffing decisions, amenity improvements, and service recovery efforts. These outcomes positioned the sentiment analysis model as a foundational component within the broader analytics pipeline.

**Model 2: Text Summarization**

The second model was developed to transform detailed guest feedback into concise, readable summaries while simultaneously extracting recurring themes across survey responses. This approach addressed two key analytical challenges: compressing lengthy textual data into digestible formats and identifying common patterns that emerged within the feedback corpus. By using advanced natural language processing techniques, the model produced summaries that preserved the core intent and emotional tone of the original comments while enhancing interpretability.

To evaluate the model's effectiveness, a subset of 100 survey entries was summarized using a few-shot prompting strategy powered by the Mistral 7B architecture. Performance was benchmarked against a fine-tuned FLAN-T5 model using standard metrics, including ROUGE and BERTScore.

**Table 1: Model Evaluations Results and Summarization Samples**

| Metric | Value |
|---|---|
| eval_loss | 45.6612 |
| eval_model_preparation_time | 0.0058 |
| eval_rouge1_f1 | 0.1586 |
| eval_rouge2_f1 | 0.0000 |
| eval_rougeL_f1 | 0.1584 |
| eval_rougeLsum_f1 | 0.1594 |
| eval_bertscore_f1 | 71.2502 |
| eval_runtime | 14.7449 |
| eval_samples_per_second | 6.7820 |
| eval_steps_per_second | 6.7820 |

As presented in Figure 6, the summarization model achieved a ROUGE-1 F1 score of 0.1586 and a ROUGE-Lsum F1 of 0.1594, with a BERTScore F1 of 71.25%, indicating reasonable semantic fidelity despite modest token-level overlap. While ROUGE-2 scores remained low, consistent with the variability in human-generated feedback, the model successfully retained critical sentiment cues and operational themes.

To supplement these metrics, a qualitative review of representative outputs was conducted.

**Table 2: Example Summaries**

| Original Feedback | Model-Generated Summary |
|---|---|
| I appreciated the fast check-in... | Positive service noted; concerns about room cleanliness and noise. |
| Everything was perfect... | Guest expressed full satisfaction with cleanliness, food, and scenery. |
| Hotel was okay but billing confusion... | Mixed experience; billing issues and missing amenities caused dissatisfaction. |

Figure 7 highlighted three illustrative examples in which original guest feedback are shown alongside their generated summaries. The selection reflected the model's ability to produce semantically aligned responses across varying sentiment classes and complexity levels.

The examples presented in Figure 7 reinforce the model's interpretive strength, showing how even highly critical feedback was distilled into structured summaries that preserved nuance, such as dissatisfaction with room conditions, poor amenities, or service-related complaints. Positive feedback was similarly refined, with emphasis maintained on features like cleanliness, food quality, or scenic views. These outputs demonstrate the model's capacity to deliver sentiment-aware summaries that can be used both for trend analysis and operational response planning.

Beyond quantitative evaluation, the summarization model was further leveraged to extract and categorize dominant themes across sentiment classes.

**Table 3: Themes**

|    | Sentiment | level_1 | Theme |
|----|-----------|---------|-------|
| 0  | Negative | Theme 1 | dissatisfied (experience) |
| 1  | Negative | Theme 2 | frustration, aroma, recognition |
| 2  | Negative | Theme 3 | overcharge |
| 3  | Negative | Theme 4 | unsanitary/dirtiness |
| 4  | Slightly Negative | Theme 1 | dirty |
| 5  | Slightly Negative | Theme 2 | disappointment (core emotion) or continental breakfast (topic) |
| 6  | Slightly Negative | Theme 3 | inconvenience |
| 7  | Slightly Negative | Theme 4 | late/cleanliness/mishandled services |
| 8  | Very Negative | Theme 1 | collision/damage |
| 9  | Very Negative | Theme 2 | disappointment |
| 10 | Very Negative | Theme 3 | dissatisfaction (with service) |
| 11 | Very Negative | Theme 4 | dissatisfied service/facilities |
| 12 | Very Negative | Theme 5 | full (parking lot) / unavailable (spaces) |
| 13 | Very Positive | Theme 1 | awesome (team) |
| 14 | Very Positive | Theme 2 | cleanliness (pool area), food options (snacks) |
| 15 | Very Positive | Theme 3 | comfort |

Figure 8 presented the results of this thematic mapping, displaying the top-ranked themes for each sentiment level derived from model-generated summaries. By structuring the feedback around recurring keywords, this figure offers a clear window into the emotional and experiential drivers behind guest perceptions.

In the negative and very negative categories, themes such as *"unsanitary/dirtiness,"* *"collision/damage,"* and *"overcharge"* consistently emerged, pointing to serious concerns around cleanliness, value perception, and operational breakdowns. The appearance of *"full (parking lot)"* and *"unavailable (spaces)"* in very negative responses suggests frustrations not just with the rooms, but with peripheral services impacting the overall stay.

Conversely, slightly negative feedback included more nuanced or mixed sentiments, with themes like *"disappointment,"* *"continental breakfast,"* and *"mishandled services"* hinting at minor service lapses rather than outright failures. Very positive responses focused instead on uplifting aspects, with guests highlighting *"awesome (team),"* *"cleanliness (pool area),"* and *"comfort"* — all of which suggest emotional satisfaction tied to ambiance and staff performance.

As illustrated in Figure 8, the stratified theme distribution by sentiment enables hotel management to link operational areas with emotional tone, transforming qualitative feedback into actionable categories. This mapping sets the stage for more precise intervention strategies, whether addressing pain points or reinforcing the experiences that guests value most.

**Model 3: Auto Replier**

The third model was constructed to facilitate automated responses to guest feedback, thereby streamlining communication and ensuring consistency in tone and messaging across varying sentiment levels. This approach aimed to assist hotel staff in crafting timely, context-aware replies while reducing the cognitive burden of manual composition. To guide the

emotional tone of the responses, a custom sentiment-to-tone mapping schema was manually defined. Each feedback entry was first classified by sentiment, which then determined the style of the reply. For example, feedback labeled as "Very Positive" was matched with an *enthusiastic and grateful* tone, whereas "Very Negative" entries received replies written in a *deeply apologetic and resolution-focused* manner. These descriptors were embedded directly into the prompt design to ensure that generated replies conveyed appropriate empathy and professionalism.

To implement the automated reply engine, the model utilized the unsloth/mistral-7b-instruct-v0.3-bnb-4bit architecture, selected for its instruction-following capability and efficient memory footprint. Configuration was handled through a lightweight wrapper designed for few-shot prompting, allowing the model to generate tailored responses with minimal overhead. The sentiment-tone relationships were encoded programmatically as a dictionary that linked each label to a specific reply style, thereby enabling controlled generation across sentiment classes. This design allowed the model to balance factual correctness with emotional intelligence, producing replies that addressed guest concerns while aligning with hospitality communication standards.

Following the technical implementation, a representative sample of auto-generated replies was extracted to assess how effectively the model personalized its responses to both content and sentiment. Figure 9 presents a structured comparison of guest feedback, assigned sentiment, mapped tone, and the corresponding reply produced by the automated system.This figure serves to illustrate the functional integration between sentiment classification and tone-controlled generation.

**Table 4: Auto Reply Examples**

| Feedback | Sentiment | Tone | Reply |
|---|---|---|---|
| The roll a way and hideaway bed mattress were terrible in this room. Hideaway was broke down Rollaway was a box with springs sticking through it. Other wise hotel and staff superb | Neutral | balanced and professional | We're truly sorry about the inconvenience with our roll-away and hideaway beds during your stay. Our quality standards fell short here, especially regarding comfort. Rest assured, we've taken note of these concerns and will ensure they're addressed promptly. Thank you for appreciating our staff and overall service—we'll do better on making your sleep experience more comfortable too! |
| This is our favorite hotel in all our travels. When we retire we plan on using the hotel much more. Thank you. | Very_Positive | professional and courteous | Delighted to hear such kind words! We appreciate your patronage immensely. Looking forward to serving you even better as retirement approaches. Warm regards. |
| The staff were awesome | Very_Positive | professional and courteous | We're thrilled to hear about your positive experience! Our team works hard to make every stay special. Thank you for recognizing their efforts. We look forward to welcoming you back soon! |
| The room was not as clean as it has been in the past. There were areas that needed dusting, and there were smears on the Windows. | Negative | apologetic and solution-focused | We're truly sorry to hear about the cleanliness issue during your stay. It seems our usual standards fell short this time around. We appreciate your detailed observation which helps us address these concerns promptly. Rest assured, we have passed along your valuable feedback to our housekeeping team who will make sure they pay special attention to window cleaning and general tidiness going forward. Thank you again for bringing this matter to our notice so we can improve. |
| Comfortable beds & Pillows - slept great. | Very_Positive | professional and courteous | Delighted to hear you had a restful sleep! Our team works hard to ensure our bedding meets high standards. We appreciate your positive experience and will share this feedback with them. Enjoy the rest of your stay! |

As shown in Figure 9, feedback labeled as "Very Positive" was met with replies that expressed sincere gratitude and recognition of the guest's praise, using concise yet emotionally

expressive language. Neutral or mixed feedback received professional and measured responses, striking a tone of balance and courtesy while acknowledging guest experiences without excessive embellishment. Negative entries prompted replies that were empathetic, direct, and solution-oriented—often referencing specific issues such as discomfort, cleanliness, or service delays with apologies and subtle corrective intent.

Each reply demonstrated tonal adherence, factual precision, and brevity suited for mobile-friendly formats, as defined by the model's prompt schema. The alignment between feedback sentiment and generated response tone in Figure 9 reinforced the model's ability to replicate emotionally intelligent communication at scale. These outputs further validated the usefulness of prompt-based tone mapping as a method for controlled, brand-consistent guest engagement.

**Conclusion**

Taken together, the three models developed in this study demonstrated a layered approach to extracting insight and enhancing guest engagement through advanced natural language processing. The sentiment classification model established a foundation for emotional understanding, accurately labeling feedback and enabling tone-specific responses. The summarization model built upon this by distilling detailed comments into interpretable formats while preserving nuance, which in turn facilitated thematic analysis across sentiment categories. Finally, the automated reply system showcased how large language models could generate empathetic, context-aware responses tailored to the emotional content of guest feedback. By integrating these components into a unified pipeline, the project provided a scalable framework for transforming unstructured survey data into actionable intelligence and brand-aligned communication. These methods offer practical utility for hospitality operations and illustrate the

broader potential of AI-driven tools in elevating customer experience through intelligent automation.

## Chapter 5: Summary

**Introduction**

Three successful models were developed through the application of deep learning and natural language processing (NLP) with hotel survey data. These included a sentiment analysis classifier designed to assess guest satisfaction, a text summarization model that distilled guest experiences into key themes, and an automated response generator that crafted replies aligned with the emotional tone of feedback. Each model contributed to a comprehensive approach for extracting and operationalizing insights from short-form textual responses.

The implementation of these models represented a comprehensive approach to addressing real-world business challenges in hospitality settings. Through the integration of advanced AI models and customized prompting strategies, the project demonstrated how advanced NLP techniques successfully managed variability in guest communication styles, demographic influences, and sentiment ambiguity. This enabled automated yet empathetic guest communication, supported service improvement strategies, and established a foundation for scalable feedback analysis.

**Summary of Project**

The project successfully fulfilled its objective of modernizing customer feedback analysis for the Hyatt Regency of Green Bay through the strategic application of Natural Language Processing (NLP) and deep learning methodologies. Utilizing a dataset comprising eight years of hotel survey responses, the study developed and implemented a systematic framework designed to enhance operational efficiency and accuracy in interpreting guest sentiment.

Three models were constructed to target key feedback-related functions: sentiment classification, text summarization, and automated reply generation. The sentiment analysis model reduced manual review workload by accurately identifying extreme emotional responses, while the text summarization model provided concise thematic overviews of guest experiences. The automated reply system enabled timely and context-aware engagement with guests based on their expressed feedback.

By integrating transformer-based language models and customizing prompting techniques, the project achieved improved responsiveness and reduced latency in addressing guest concerns. The approach demonstrated practical viability in a hospitality context and established a scalable foundation for future adoption across similar service-oriented environments. Overall, the project met its goals by replacing manual processes with intelligent automation and enhancing both the interpretability and actionability of guest feedback.

**Implications of Project Results**

By demonstrating the viability of deep learning and natural language processing (NLP) techniques for analyzing hotel survey data, the project reinforced the value of intelligent automation in enhancing operational efficiency and customer engagement. The sentiment analysis model enabled prioritization towards feedback based on emotional tone, allowing staff to allocate attention more effectively to guests expressing dissatisfaction. The summarization model provided streamlined insights into common themes and concerns, improving managerial decision-making and response planning. Additionally, the automated reply generator offered a scalable solution for timely guest communication, contributing to stronger guest satisfaction and loyalty metrics.

These findings were not limited to the hotel industry; rather, they demonstrated potential for broader application across other sectors that regularly manage large volumes of unstructured customer feedback. Industries such as retail, healthcare, education, and transportation could similarly benefit from NLP-driven tools to streamline feedback interpretation, support automated engagement, and uncover actionable insights from freeform text data. By integrating such tools, organizations can gain a competitive advantage through enhanced responsiveness, reduced processing time, and more informed decision-making based on real-time customer insights.

**Limitations and Next Steps**

While the project had achieved its stated objectives, several limitations had emerged that constrained model performance and broader experimentation. One key limitation involved limited familiarity with API integration and access management, particularly regarding paid services. Due to budgetary concerns and uncertainty about their potential impact on model accuracy, advanced APIs had not been utilized, thereby restricting experimentation with potentially higher-performing third-party tools.

Additionally, the sentiment analysis model had reached an accuracy of approximately 65%, which fell below the targeted threshold of 80%. Although the model had demonstrated reliability in identifying feedback that was strongly positive or strongly negative, it had exhibited lower precision when classifying responses that were neutral, mixed, or contained subtle tonal shifts. This limitation had underscored the need for a more granular labeling approach. In future iterations, manual annotation of the dataset—using human interpretation across nuanced sentiment categories such as "Very Positive," "Positive but Critical," "Mixed/Neutral," "Negative," and "Very Negative"—would have strengthened the training data and improved classification accuracy. Achieving an accuracy rate of 90% or higher would have supported

broader application of the sentiment model across other survey datasets, enhancing its value as a scalable tool for intelligent feedback analysis in various industries.

Although the summarization and auto-reply models had performed well overall, occasional limitations had been observed in generating context-specific summaries and emotionally attuned responses. Further tuning of prompt templates and expansion of training examples may have improved their contextual sensitivity and relevance. Additionally, incorporating user feedback mechanisms—such as staff evaluations or guest satisfaction tracking—could have helped identify tone mismatches and guided further refinement. Future development may also have benefited from techniques such as reinforcement learning or sentiment-aligned scoring functions to more closely align model outputs with business objectives and emotional nuance.

**Conclusion**

This project successfully demonstrated the potential of data science, specifically deep learning and natural language processing (NLP), to improve customer feedback systems in the hospitality industry. Through the development of three integrated models for sentiment analysis, text summarization, and automated reply generation, the study addressed key operational challenges faced by the Hyatt Regency of Green Bay. These models enabled faster interpretation of survey feedback, reduced manual review burden, and supported more timely and personalized guest communication.

The application of advanced AI tools to interpret how guests expressed their thoughts and emotions had provided a valuable foundation for innovation within the hotel industry. Despite limitations in model accuracy and restricted access to certain technologies, the system had

delivered meaningful insights and had demonstrated strong potential for scalable solutions capable of being adapted across multiple hotel properties.

From a business perspective, this work offered a proof of concept for integrating AI-driven solutions into service workflows. By prioritizing sentiment, identifying recurring themes, and automating guest outreach, hotels could improve response efficiency, enhance guest satisfaction, and inform strategic decision-making. More broadly, the project reinforced the value of applied data science in converting unstructured textual feedback into actionable intelligence—an approach that holds promise across multiple service-oriented industries seeking to optimize customer engagement.

**References**

Adil, M., Ansari, M. F., Alahmadi, A., Wu, J. Z., & Chakrabortty, R. K. (2021). Solving the
problem of class imbalance in the prediction of hotel cancelations: A hybridized machine
learning approach. *Processes*, *9*(10), 1713.

Alotaibi, E. (2020). Application of machine learning in the hotel industry: a critical review.
*Journal of Association of Arab Universities for Tourism and Hospitality*, *18*(3), 78-96.

Bompotas, A., Ilias, A., Kanavos, A., Makris, C., Rompolas, G., & Savvopoulos, A. (2020). A
sentiment-based hotel review summarization using machine learning techniques. In
*Artificial Intelligence Applications and Innovations. AIAI 2020 IFIP WG 12.5
International Workshops: MHDW 2020 and 5G-PINE 2020, Neos Marmaras, Greece,
June 5–7, 2020, Proceedings 16* (pp. 155-164). Springer International Publishing.

Bonny, Jannatul Jahan, et al. "Deep learning approach for sentimental analysis of hotel review on
bengali text." *2022 Second International Conference on Advances in Electrical,
Computing, Communication and Sustainable Technologies (ICAECT)*. IEEE, 2022.

Chang, V., Liu, L., Xu, Q., Li, T., & Hsu, C. H. (2023). An improved model for sentiment
analysis on luxury hotel review. *Expert Systems*, *40*(2), e12580.

Hossen, M. S., Jony, A. H., Tabassum, T., Islam, M. T., Rahman, M. M., & Khatun, T. (2021,
March). Hotel review analysis for the prediction of business using deep learning
approach. In *2021 international conference on artificial intelligence and smart systems
(ICAIS)* (pp. 1489-1494). IEEE.

Pang, T., Liu, J., Han, L., Liu, H., & Yan, D. (2025). A Deep Learning-Based Analysis of

    Customer Concerns and Satisfaction: Enhancing Sustainable Practices in Luxury Hotels.

    *Sustainability*, *17*(8), 3603.

Popovic, O. B., Nikic, V., Bulatovic, I., & Delibasic, M. (2018). Modeling perceived quality,

    customer satisfaction and probability of guest returning to the destination. *Montenegrin*

    *Journal of Economics*, *14*(1), 69-78.

Ku, C. H., Chang, Y. C., Wang, Y., Chen, C. H., & Hsiao, S. H. (2019). Artificial intelligence and

    visual analytics: a deep-learning approach to analyze hotel reviews & responses.

Sanwal, M., & Mazhar, M. M. (2023). Performance Comparison of Machine Learning and Deep

    Learning Models for Sentiment Analysis of Hotel Reviews. *Int. J. Inf. Tec. App. Sci*, *5*(1).

Sim, J., Mak, B., & Jones, D. (2006). A model of customer satisfaction and retention for hotels.

    *Journal of Quality Assurance in Hospitality & Tourism*, *7*(3), 1-23.

Worsfold, K., Fisher, R., McPhail, R., Francis, M., & Thomas, A. (2016). Satisfaction, value and

    intention to return in hotels. *International Journal of Contemporary Hospitality*

    *Management*, *28*(11), 2570-2588.

**Appendix**

The code for this project can be found at the following location:

https://github.com/caelisaa/TEST/tree/master/CAPSTONE/notebooks