Chapter 4: Deep Survey Analysis

**Introduction**

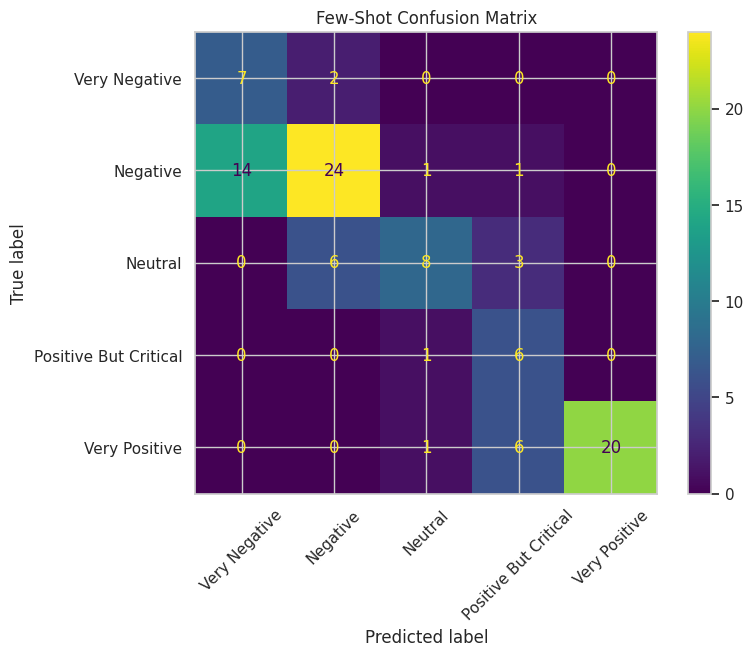
The hospitality industry had long struggled with extracting meaningful insights from guest survey feedback due to the time-intensive nature of manually reviewing open-ended comments. This process often resulted in delayed recognition of operational issues, missed opportunities for service improvement, 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 5 Few-Shot Confusion Matrix**

****

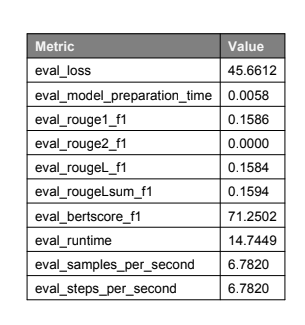
The results presented in Figure 5 reinforced the model’s suitability for operational deployment, particularly in its ability to identify satisfaction extremes that warranted immediate attention. For example, the high recall associated with “Very Negative” feedback enabled hotel staff to prioritize and 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.

**Figure 6: Model Evaluation Results and Summarization Samples**

****

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.

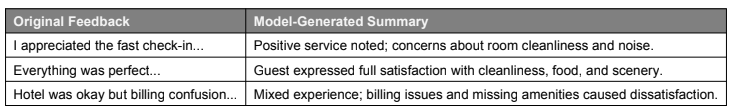
**Figure 7: Example Summaries**

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.

**Figure 8: Themes**

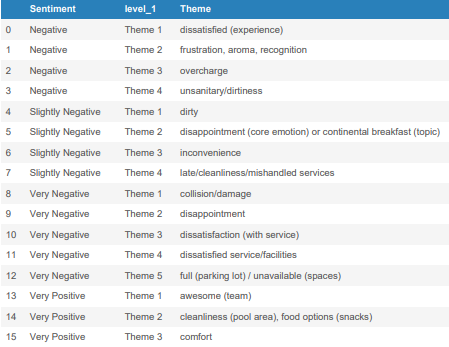


Figure 8 presented the results of this thematic mapping, displaying the top-ranked themes for each sentiment level derived from model-generated summaries. By structuring the feedback around recurring keywords, this figure offers a clear window into the emotional and experiential drivers behind guest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

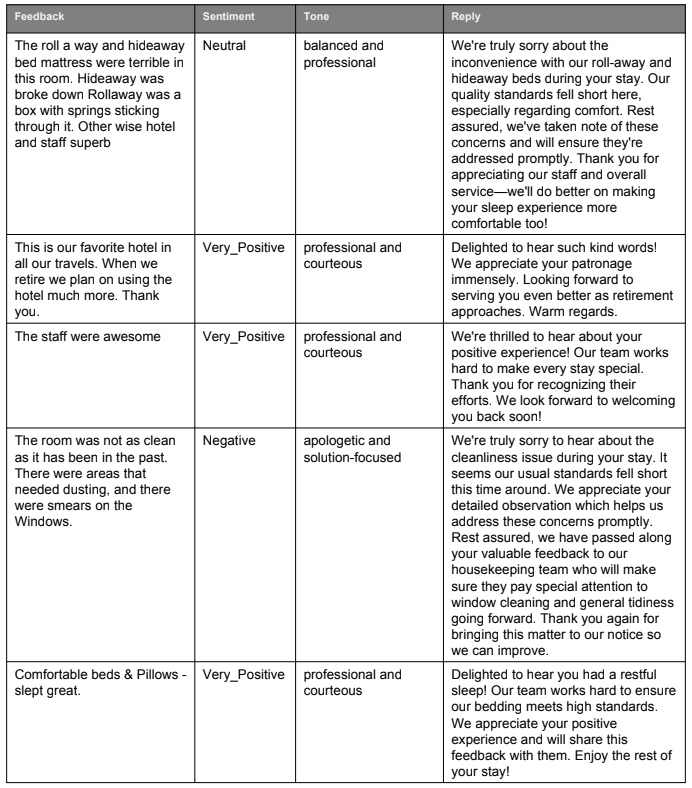
**Model 3: Auto Replier**

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.

**Figure 9: Auto Reply Examples**

****

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.