

# Using Multitask Learning with Pre-trained Language Models for Aspect-Based Sentiment Analysis in the Hospitality Industry

Xuan-Yu You<sup>1</sup>, Shih-Chuan Chang<sup>1</sup>, Sheng-Mao Hung<sup>1</sup>,  
Chih-Hao Ku<sup>2</sup>, Yung-Chun Chang<sup>1\*</sup>

<sup>1</sup>Graduate Institute of Data Science, Taipei Medical University, Taiwan

<sup>2</sup>Department of Information Technology and Decision Science, University of North Texas, USA

## Abstract

With the rise of the Internet, online reviews have become crucial for consumer purchase decisions, as they often contain valuable insights into user experiences. Despite the abundance of user-generated data on social media and other platforms, it remains largely underutilized. This study enhances sentiment analysis of online hotel reviews by employing Pre-trained Language Models (PLMs) such as BERT, RoBERTa, and ALBERT, which significantly outperform traditional methods in capturing textual nuances. Our comparative analysis shows that RoBERTa excels, achieving the highest ROC AUC of 0.8717 and AUPRC of 0.7895 for predicting travel types and an AUC of 0.9218 with an AUPRC of 0.6521 for sentiment analysis. Results highlight varied sentiment expressions among different traveler types, with business travelers typically more critical. These insights contribute to academic research and empower hotel managers to tailor services and improve guest experiences based on detailed feedback from customer reviews.

## 1 Introduction

Since the invention of the Internet, the sharp increase of online platforms such as TripAdvisor and Booking.com has revolutionized the landscape of consumer feedback in the hospitality industry. These platforms offer media for users to share their subjective opinions, recommendations, and ratings on their accommodation experiences. This tendency profoundly impacts hotels' reputational dynamics and managerial strategies (Abrahams et al., 2015). TripAdvisor, the largest travel platform (Yu, et al., 2017) alone, amasses over 600 million reviews and opinions, highlighting its prominent role in shaping consumer behavior and business

strategies. Whereas such democratization of customer feedback allowed consumers to gather information efficiently, this trend simultaneously introduces complex analytical challenges due to the nuanced sentiments embedded in hotel guests' rich, multifaceted data. Traditional sentiment analysis methodologies often fall short when addressing the multiaxial and contextually rich data that modern hotel reviews represent. These methodologies typically simplify sentiments into binary positive and negative dichotomies, which are insufficient for capturing the subtleties required in the hospitality context (De Pelsmacker et al., 2018; Gavilan et al., 2018; Hernández-Ortega, 2018).

To address this challenge, this research leverages recent advancements in artificial intelligence, specifically deep learning technologies. These technologies have introduced intricate models of Pre-trained Language Models (PLMs) — BERT, RoBERTa, and ALBERT — which demonstrate an enhanced capacity for understanding and processing human language. These models utilize extensive pre-trained contextual embeddings, allowing deeper and more accurate classification of textual data based on sentiment and thematic depth. This marks a significant improvement over earlier models, such as TextCNN and LSTM-ATT, which capture local features and sequential information but lack the depth provided by PLMs (Zhao et al., 2019). PLMs are theorized to enhance the analysis of hotel reviews through a multi-dimensional approach. The goal is to create an advanced analytical model that predicts overall sentiment and delves into the complex aspects of service quality, cleanliness, location, and value. These factors are crucial for shaping business strategies and improving customer satisfaction in the hospitality industry, demanding the sophisticated use of BERT, RoBERTa, and ALBERT to transform customer feedback management and elevate both guest experiences and operational efficiencies (Eivind et al., 2012; Filieri et al., 2015; Jin et al., 2017; Schuckert et al., 2015). The performance of PLM frameworks will be evaluated against traditional machine learning methods such as Naive Bayes, Random Forest, and XGBoost, as well as other neural networks such as LSTM-ATT, MLP, and TextCNN. This comparison

aims to establish benchmarks for their real-world effectiveness. Additionally, the project will develop a comprehensive strategy for integrating insights from PLM analysis into practical hotel management practices. It will also rigorously assess the impacts of these advanced PLM applications on hotel management decision-making processes, focusing on customer satisfaction and overall business performance.

This research employs advanced natural language processing (NLP) techniques to enhance the comprehension of customer sentiments, thereby providing hotel managers with data-driven strategies to improve service quality. The expected outcomes of this study are poised for managing customer feedback in the hospitality industry through the implementation of progressive PLM technology. This approach contributes significantly to both academic research and practical applications by enabling industry professionals to leverage big data and analytical tools effectively to optimize service delivery and customer satisfaction.

## 2 Related Work

### 2.1 Online Reviews and Ratings in the Hospitality Sector

The significant influence of online hotel reviews on consumer behavior is well-recognized, underscoring their importance in digital tourism and hospitality. Travelers depend on electronic word-of-mouth (eWOM) from platforms like TripAdvisor, which impacts purchase decisions, revisit intentions, and satisfaction (Mauri et al., 2013; Ögüt et al., 2012). These platforms aggregate ratings that influence bookings and perceptions of service quality (Noone et al., 2015; Schuckert et al., 2015) and allow exploration of how managerial responses improve customer relationships and business performance (Wang et al., 2018; Xie et al., 2014). The dynamic between consumer feedback and business response is crucial, with both positive and negative reviews affecting consumer loyalty and purchase intentions, particularly when businesses engage with reviews (Zhao et al., 2019). Social media analysis offers insights into user sentiments, highlighting these platforms' role in business strategy (Lu et al., 2015; Herrero et al., 2015). Reviews, providing both ratings and qualitative feedback, shape customer expectations and decisions, which are essential for marketers using sentiment analysis to enhance services (Huang et al., 2013). Thus, strategic use of online reviews is

vital for any hospitality business aiming to thrive in a competitive environment.

### 2.2 Aspect-Based Sentiment Analysis

Aspect-based sentiment analysis (ABSA) is a significant advancement in sentiment analysis, focusing on the precise sentiments associated with specific aspects of products or services rather than overall sentiment. This approach is especially relevant in sectors like hospitality, where feedback can vary widely across aspects like cleanliness, location, or staff behavior (Cambria et al., 2017; Hu et al., 2017). ABSA has evolved from rule-based systems to machine learning techniques, including supervised learning that utilizes labeled data to classify aspects and sentiments (Schouten et al., 2016). Techniques like Latent Dirichlet Allocation (LDA) have been used for topic modeling to uncover latent aspects within datasets (Blei et al., 2003). Additionally, advanced models such as conditional random fields (CRFs) and graph-based co-ranking algorithms leverage syntactic and semantic relationships to enhance the extraction and ranking of aspect-related sentiments (Jakob et al., 2010; Liu et al., 2015).

Recent innovations in ABSA include structural topic models and sentiment-sensitive frameworks that consider both the content and context of reviews, offering deeper insights into consumer behavior and service quality (Korfiatis et al., 2019). The use of these sophisticated techniques has shown potential in improving service customization and operational efficiency, indicating the importance of context and granularity in sentiment analysis (Chang et al., 2019; Sann et al., 2020). By applying these advanced methods, businesses can derive actionable insights crucial for enhancing customer satisfaction and maintaining a competitive edge.

## 3 MultiTask PLMs for Prediction of Travel Types and Aspect-Based Sentiment Analysis

Considering the potential correlation between travel type and aspect-sentiment, we adopted a multitask learning framework as the primary architecture for our model. As depicted in Figure 1, the proposed architecture employs PLMs configured for multitask learning, enabling simultaneous processing of both Travel Type Prediction (TTP) and ABSA tasks. The architecture

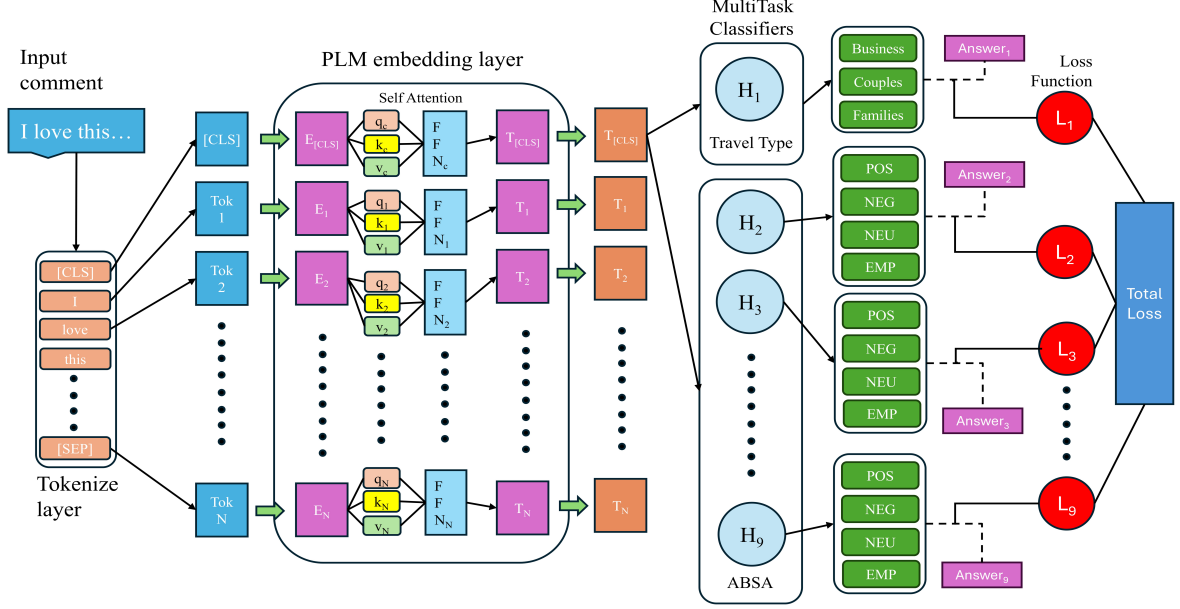


Figure 1: MultiTask PLMs for Prediction of Travel Type and ABSA Joint Learning Structure Plot.

begins with a Tokenization layer, which performs the initial tokenization of the input text. This is followed by the PLM Embedding layer, which is responsible for generating comprehensive text representations. In this research, we evaluate the performance of three distinct PLMs, specifically Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018), the robustly optimized BERT pretraining approach (RoBERTa) (Liu et al., 2019), and A Lite BERT (ALBERT) (Lan et al., 2019). These models represent significant advancements in the field of natural language processing, and our research aims to discern their effectiveness across various computational tasks. Subsequently, the MultiTask Classifiers are utilized to perform the learning tasks for each classifier involved in the model. Finally, a unified Loss Function computes the loss for each task, serving as the foundation for the multitask learning approach. This integrated architecture ensures efficient learning and improved performance across both tasks, leveraging the inherent synergies between travel type classification and sentiment analysis.

### 3.1 Tokenize Layer

In our study, we utilize the default tokenizer pretrained for three PLMs. The primary function of the Tokenization layer is to transform raw text into a structured format that is comprehensible by the model. Initially, this layer conducts basic tokenization by segmenting the text into individual words and symbols. For models such as ours that

implement subword tokenization, this stage further decomposes words into smaller, more manageable subunits. Each token is then assigned a unique identifier from a pre-established dictionary. Moreover, several special symbols are incorporated to enhance the model's understanding and processing capabilities. These include the [CLS] symbol, which is positioned at the beginning of each sentence to signify the start; the [SEP] symbol, used to demarcate separate sentences within the same input; the [PAD] symbol, which standardizes the lengths of inputs for batch processing; and the [MASK] symbol, employed to obscure certain tokens randomly during the training phase to prevent the model from merely memorizing the data. To ensure uniformity in processing, all input sequences are adjusted to a consistent length. This standardization is crucial for efficient batch processing and facilitates the model's ability to learn from and make predictions based on the input data effectively.

### 3.2 PLM Embedding Layer

The PLM embedding layer is instrumental in converting the discrete tokens generated by the tokenization layer into dense vector representations, known as embeddings. These embeddings are engineered to encapsulate both the semantic attributes and contextual nuances of words, facilitating a deeper understanding of textual data. In our PLM architecture, the embedding process is comprehensive, involving several components. Primarily, it integrates word embeddings that capture lexical semantics. Concurrently, positional embeddings are incorporated to encode the relative positions of

tokens within sentences, thereby preserving the syntactic structure of the text. Additionally, segment embeddings are utilized to differentiate between various sentences or paragraphs, ensuring that the model maintains contextual awareness across different segments of text.

Formally, if  $x_i$  denotes a token, the embedding layer transforms  $x_i$  into a high-dimensional space using the embedding function  $E$ , resulting in a vector  $v_i = E(x_i)$ . This vector is then augmented with positional and segment information to produce a comprehensive representation where  $p(pos_i)$  and  $S(seg_i)$  represent the positional and segment embeddings corresponding to the token's position  $p(pos_i)$  and segment  $S(seg_i)$ , respectively. These enriched vector embeddings  $v'_i$  are subsequently employed as input features for multitask classifiers. These classifiers are specifically designed to handle complex NLP tasks, such as determining travel types and performing ABSA. By processing these sophisticated inputs, the classifiers can more accurately predict outcomes across varied linguistic contexts. Therefore, the PLM embedding layer plays a pivotal role in transforming raw textual data into a structured format that is amenable to machine processing. This transformation is crucial for enabling the model to conduct a deep semantic analysis and extensive contextual evaluation of the text, thereby significantly enhancing the model's versatility and effectiveness in managing diverse language-based tasks.

### 3.3 MultiTask Classifiers

The MultiTask Classifiers in our architecture exploit the [CLS] token output from the PLM embedding layer, which is specifically designed to encapsulate the overall context of the input text. This classifier harnesses these comprehensive embeddings to efficiently execute multiple NLP tasks concurrently. Within the classifier, each task-specific layer is linear-based and utilizes the [CLS] token as a focal point for extracting and synthesizing a holistic understanding of the text. This mechanism enables the classifier to make nuanced and specialized predictions across various domains, such as identifying travel types and conducting ABSA.

Mathematically, the classifier can be described as follows: let  $e_{CLS}$  represent the embedding of the [CLS] token, the task-specific layer for the  $k$ -th task processes  $e_{CLS}$  to predict the outcome  $y_k = T_k(e_{CLS})$ , where  $T_k$  is typically a linear transformation followed by a non-linear activation function tailored to the specifics of each task. This

multitask learning approach not only amplifies the efficiency of the training process by leveraging shared features across different tasks but also significantly enhances the model's capacity to generalize. By sharing a common representation, the model minimizes the risk of overfitting to a specific task and maintains a high degree of adaptability, thereby improving its performance and flexibility when faced with new or evolving challenges. This methodological framework positions our model at the forefront of current NLP applications, optimizing both performance and scalability.

### 3.4 Loss Function

The loss function for our multi-task learning model is designed to simultaneously accommodate nine different tasks, with each task contributing equally to the overall loss. This is achieved by summing the individual cross-entropy losses associated with each task. Mathematically, the loss function  $L$  is represented as follows:

$$L = \sum_{i=1}^9 \text{CrossEntropy}(y_i, \hat{y}_i) \quad (1)$$

Here,  $y_i$  and  $\hat{y}_i$  represent the true and predicted values for each task respectively. The index  $i$  includes one task for TTP and eight distinct types of ABSA. This formulation ensures that the model is optimized for performance across all tasks by minimizing the prediction error uniformly across the different domains.

## 4 Experiment

In our dataset, we combined several key data as inputs for our models, including the *Review's Star Rating*, *Review's Content*, *Tourist's Travel Style*, *Trip Collective Total Points*, and *Address*. These pieces of information are concatenated with commas (","), After such preprocessing, the text data is fed into our models to predict the tourist's Travel Type and their Aspect-Sentiment related to various aspects of the trip. Travel Type includes three categories: *Business*, *Couples*, and *Families*, while Aspect-Sentiment is divided into eight categories, including *Sleep Quality*, *Location*, *Value*, *Cleanliness*, *Service*, *Business Service*, *Check-in*, and *Rooms*, each with four possible emotional states: Positive (POS), Negative (NEG), Neutral (NEU), and Empty (EMP). The Empty label was kept intentionally to better represent reality, as customers are prone to reflect on the aspects they are particularly interested in rather than the entire eight aspects. This design allows the model to capture and predict the travelers'

emotional responses in detail. The dataset is split into training, validation, and test sets in a 3:2:2 ratio. All models are first fine-tuned using the training dataset and then evaluated using the test data. Our evaluation strategies include macro average F<sub>1</sub>-score, Precision, Recall, Area Under the Receiver Operating Characteristic Curve (AUROC), and Area Under the Precision-Recall Curve (AUPRC). Precision measures the accuracy of review predictions, while recall assesses the model’s ability to identify review types. When categories are unbalanced, and we wish the predictive effects of all categories to be equally important, the macro average F<sub>1</sub>-score is an ideal metric. It balances the influence of each category, preventing it from being overshadowed by some categories’ high precision or recall rates. AUROC is suitable for evaluating model performance with balanced data, and AUPRC is suitable for evaluating model performance with unbalanced data.

#### 4.1 Dataset

To test our method, we used the dataset that we had gathered. The Hilton Hotel was chosen as our subject for this research for a few reasons. Firstly, according to Brand Finance, the global consultancy firm that specializes in brand valuation, Hilton was the most renowned and valuable brand in the industry, with a value of US \$7.8 billion in 2016, and it has secured its reign in the hospitality industry for consecutive years while overall sector growth slows. Secondly, we implemented a comparison test utilizing Google Trends across a wide range of hotel brands, including Marriott, Hilton, Holiday Inn, Hyatt, Sheraton, etc. and discovered that “Hilton” is the most frequently searched keyword among all hotel brands. For the platform, we selected TripAdvisor as it is widely recognized by travelers around the globe, and additionally, it offers an immense volume of user-generated content. Choosing a highly trafficked agency such as TripAdvisor significantly enhances our chances of gathering abundant data that is rich in detail. Further, more than a simple overall rating, reviewers can easily rate eight additional aspects of Value, Location, Sleep Quality, Rooms, Cleanliness, Service, Check-in, and Business Service. These ratings ranged from 1 to 5 stars, providing a solid basis for quantitative assessments of our approach. Customer profiles and hotel features such as Location, Highlight, and Amenities were collected as well.

Subsequent to data retrieval, basic preprocessing was undertaken to prepare the

dataset for analysis. This process involved defining the three classes of travel types, namely business, couple, and family, and the sentiment of customer reviews based on their star ratings. Specifically, reviews were categorized as follows: ratings of 1 to 2 stars were labeled as negative, a 3-star rating was considered neutral, and ratings of 4 to 5 stars were classified as positive. For the eight aspects, in addition to positive, neutral, and negative, another “empty” label is defined as aforementioned in the Experiment section. This categorization facilitates a structured approach to sentiment analysis, allowing for a detailed understanding of consumer perceptions across a spectrum of feedback.

After such a process, our dataset contains 70,000 reviews from 749 Hilton hotels in the U.S. As for the characteristics of the dataset, the distribution of travel types is even, respectively accounting for 30 to 35%. The eight sentiments, however, show the imbalance nature. The travel type distribution is as follows: Business travelers constitute 35.92% with 251,141 individuals, couples make up 32.73% with 22,909 individuals, and families represent 31.36% with 21,950 individuals. The data distribution of ABSA is shown in Figure 2.

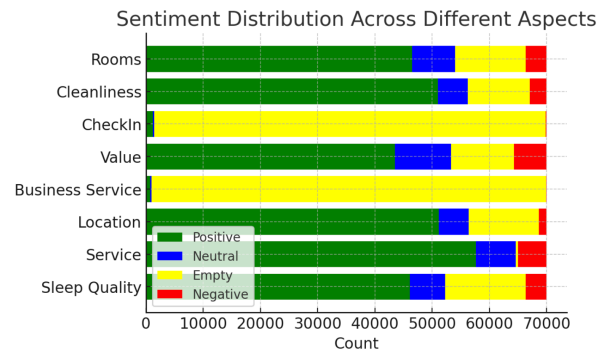


Figure 2: Data distribution of ABSA.

#### 4.2 Experiment Setup

In this study, we conduct a comparative analysis of Pre-trained Language Models (PLMs), machine learning techniques, and deep neural networks in text processing applications. For the machine learning approach, we employ TF-IDF for text embedding and feature extraction. The embedded texts are utilized by two specific models: the Travel Type model and the Aspect-Sentiment model, which are designed to predict travel classifications and multi-dimensional sentiments, respectively. We implement three widely-used machine learning algorithms—Naïve Bayes (NB), Random Forest (RF), and eXtreme Gradient



Models	Travel Type Prediction (TTP)	Aspect-Based Sentiment Analysis (ABSA)
	<i>Precision / Recall / F<sub>1</sub>-score / AUROC / AUPRC</i>	
NB	0.6113 / 0.6045 / 0.6009 / 0.7896 / 0.6589	0.4398 / 0.2617 / 0.2399 / 0.7027 / 0.3990
RF	0.6524 / 0.6453 / 0.6480 / 0.8210 / 0.7056	0.5197 / 0.3745 / 0.3626 / 0.8726 / 0.5160
XGBoost	0.6717 / 0.6657 / 0.6670 / 0.8414 / 0.7399	0.5620 / 0.5088 / 0.4938 / 0.9109 / 0.5665
MLP	0.6422 / 0.6417 / 0.6418 / 0.8239 / 0.7169	0.3418 / 0.2466 / 0.2657 / 0.7755 / 0.3864
TextCNN	0.6200 / 0.6201 / 0.6203 / 0.8043 / 0.6844	0.5327 / 0.5013 / 0.4892 / 0.8849 / 0.5268
LSTM-Att	0.6651 / 0.6621 / 0.6634 / 0.8391 / 0.7372	0.5299 / 0.4439 / 0.4208 / 0.8965 / 0.5288
ALBERT	0.6974 / 0.6928 / 0.6933 / 0.8657 / 0.7816	0.5951 / 0.5418 / 0.5421 / 0.9172 / 0.5973
BERT	0.6952 / 0.6942 / 0.6946 / 0.8678 / 0.7835	0.6076 / 0.5529 / 0.5522 / 0.9188 / 0.6079
RoBERTa	<b>0.7082 / 0.7010 / 0.7018 / 0.8718 / 0.7895</b>	<b>0.6169 / 0.5819 / 0.5817 / 0.9214 / 0.6152</b>

Table 1: Comparative Performance Metrics of Various Models for TTP and ABSA.

Boosting (XGBoost)—to train these models and optimize their performance.

In our deep learning approach, we similarly use TF-IDF for text embedding. The processed data are then input into three neural network architectures: a Multi-Layer Perceptron (MLP), a Text Convolutional Neural Network (TextCNN), and an LSTM with Self-Attention (LSTM-ATT). The MLP model comprises two fully connected layers followed by a linear classifier. The TextCNN model processes inputs through a convolutional layer before passing them through two fully connected layers. The LSTM-ATT model features a bidirectional LSTM layer to discern complex textual relationships, augmented by a self-attention layer that prioritizes significant features while diminishing the less relevant ones. This enhanced data is finally projected through a fully connected layer to produce precise output predictions.

Furthermore, we evaluated the efficacy of these models under uniform experimental conditions. All models were trained with a batch size of 16 to balance the computational load and memory usage. Specific learning rates were set—0.00001 for the PLM and 0.00005 for the other models—to foster quick convergence. We utilized the Adam optimizer for its robustness in managing sparse gradients, which is common in text data applications. To mitigate overfitting, an early stopping protocol was enforced, terminating training if no improvement in validation loss was detected after two epochs. The models utilized the cross-entropy loss function, which is suitable for the classification tasks at hand. Each model's hidden dimensions were tailored—768 for the PLM, 128 for LSTM-ATT, 64 for MLP, and 256 for TextCNN—to optimize their text processing capabilities. These configurations were based on preliminary experiments and a review of the literature, ensuring a rigorous and fair comparison

of each model's performance in handling textual data.

### 4.3 Results and Discussion

The comparative analysis of various models for predicting travel type and average aspect sentiment type reveals significant performance differences, particularly highlighting the superiority of RoBERTa. As shown in Table 1 and Fig. 3, RoBERTa consistently outperforms other models with the highest AUC and AUPRC values for both travel type (ROC AUC = 0.8717, AUPRC = 0.7895) and sentiment analysis (AUC = 0.9218, AUPRC = 0.6521). This demonstrates its robustness in handling both balanced and imbalanced datasets. In contrast, LSTM-Att and XGBoost, although performing well, fall behind in multi-task settings. Traditional machine learning models like Random Forest and Naive Bayes exhibit considerably lower AUC and AUPRC values, underscoring their limitations in complex semantic parsing and sentiment analysis tasks. The results underscore the pivotal role of advanced NLP techniques in enhancing model accuracy in the hospitality industry. Specifically, the superior performance of BERT-based models like RoBERTa suggests that contextually aware language models can significantly improve the extraction and classification of nuanced sentiment from customer reviews, leading to more informed decision-making and improved customer satisfaction in hospitality management.

From the performance of the models, it is notable that the MLP performs differently on the ABSA task compared to the Travel Type task, with its F<sub>1</sub>-score being lower than traditional machine learning methods and closely matching that of the most basic Naive Bayes. This phenomenon may be attributed to limitations in the training samples, particularly in certain extremely unbalanced sub-tasks, which, in turn, may have led to overfitting

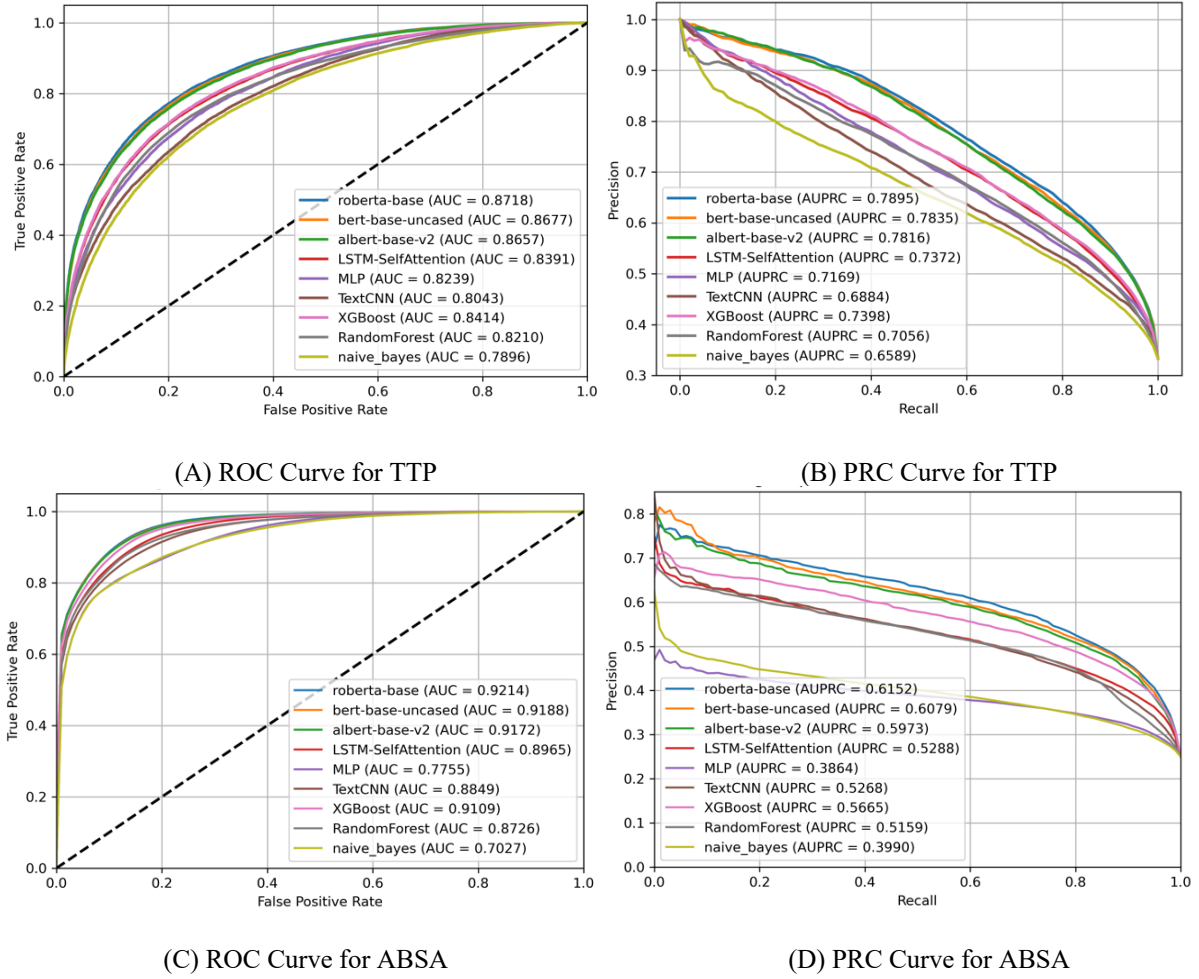


Figure 3: ROC and PRC Curves for Travel Type Prediction and Aspect-Based Sentiment Analysis.

in the MLP model. In our data visualization, it is observed that most of the reviews are concentrated on "Service," while feedback on "Business Service" and "Check-In" is relatively scarce. Such a distribution of data typically results in many positive reviews. Through a thorough analysis and visualization of the review data, we have uncovered some potential insights. If these review trends can be accurately predicted, it would not only help hotel operators avoid potential survivorship bias but also enable them to make beneficial improvements based on negative feedback. This not only helps operators better understand customer needs but is also an important step towards achieving sustainable business objectives.

Taking four ABSA subtasks as examples and through the visualization results in Figure 3, we gain a deeper understanding of the distribution of hotel service evaluations by different travel types. The charts show that business travelers tend to provide neutral or negative feedback on sleep quality, possibly reflecting a shortfall in meeting

the needs of this traveler segment. Meanwhile, couples often leave a higher proportion of blank evaluations, which might indicate a less proactive approach to reviewing experiences that need to meet good or bad standards. Additionally, regardless of travel type, there is generally enthusiastic participation in evaluating service quality, with a significantly higher proportion of positive feedback than other aspects. This suggests that the overall service quality of the hotels generally receives approval from guests. However, most of the feedback on value tends towards neutral to negative, which may imply that the hotels need to improve their cost-effectiveness. Such visualized data not only reveals the overall satisfaction levels of guests but also guides hotel management on which areas need improvement to enhance the customer experience.

Our model demonstrates significant potential, surpassing traditional deep learning and machine learning methods that do not utilize multitask learning. Notably, the RoBERTa model achieves not only higher precision and recall than other

models, but the gap between these metrics is also remarkably close. This performance allows our model to excel in scenarios with unbalanced data, such as the ABSA task, where it effectively captures minority classes. This may be attributed to RoBERTa being trained on a substantial amount of data during the pre-training phase, which was extended further, and its dynamic adjustment of the masking pattern during training. This enhancement enables the model to perform exceptionally well in specific tasks, such as understanding diverse customer sentiments and preferences, which are often embedded in unstructured text data. By accurately classifying and predicting travel types and sentiment aspects, the RoBERTa not only enhances the precision of data analysis but also contributes to more informed strategic decision-making within the hospitality sector.

Furthermore, the study underscores the challenge for traditional machine learning models in keeping pace with the depth and variability of data that modern NLP tasks demand. While models like Random Forest and Naive Bayes show resilience in simpler tasks, their performance significantly drops in multi-faceted sentiment analysis, indicating a need for more robust, adaptable algorithms that can handle the complexities of real-world data. Incorporating BERT-based models into practical applications could revolutionize customer relationship management by providing insights that enable personalized customer interactions and proactive service adjustments. This strategic integration of NLP technologies promises not only to elevate customer satisfaction and loyalty but also to drive business growth through more nuanced engagement strategies.

## 5 Conclusion

In conclusion, this study highlights the effectiveness and necessity of advanced NLP techniques, particularly BERT-based models of RoBERTa, in the hospitality industry. By evaluating the performance of multiple models on travel type prediction and sentiment analysis tasks, the research demonstrates that RoBERTa's robust handling of both balanced and imbalanced data yields superior results, particularly in capturing nuanced sentiment aspects critical for strategic decision-making. The study's findings underscore that traditional machine learning models, though effective for simpler tasks, fall short in handling the complexity of real-world, unstructured data found in customer reviews.

The research aimed to improve model interpretability and application in the hotel industry, which was achieved through analysis of aspect-based sentiment (ABSA) across different travel types. These insights reveal critical trends, such as business travelers' concerns with sleep quality and a consistent emphasis on service quality among guests. By pinpointing areas like value perception that need improvement, this study offers actionable insights for hotel operators to refine customer experience strategies.

These results suggest that AI models trained to parse intricate sentiment can serve as essential tools in customer relationship management, enabling hotels to personalize guest interactions and make data-driven improvements. Future work could build on these insights by exploring hybrid models that combine traditional and neural network approaches, enhancing both model efficiency and predictive accuracy. As AI continues to evolve, integrating such models in hospitality has the potential to redefine service excellence and foster sustained business growth.

## Acknowledgments

This study was supported by the National Science and Technology Council of Taiwan under grants NSTC 113-2627-M-006-005-, NSTC 113-2221-E-038-019-MY3, and National Health Research Institutes under grants NHRI-13A1-PHCO-1823244.

## References

- Abrahams, A. S., Fan, W., Wang, G. A., Zhang, Z., & Jiao, J. 2015. [An integrated text analytic framework for product defect discovery](https://journals.sagepub.com/doi/abs/10.1111/poms.12303). *Production and Operations Management*, 24(6), 975-990. <https://journals.sagepub.com/doi/abs/10.1111/poms.12303>.
- Blei, D. M., Ng, A. Y., Jordan, M. I. 2003. [Latent dirichlet allocation](https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?ref=http://githubhelp.com). *The Journal of Machine Learning Research*, 3, pp 993–1022.
- Association for Computing Machinery. 1983. *Computing Reviews*, 24(11), 503-512. <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?ref=http://githubhelp.com>.
- Cambria, E., Poria, S., Gelbukh, A., Thelwall, M. 2017. [Sentiment Analysis Is a Big Suitcase](https://ieeexplore.ieee.org/document/8267597). *IEEE Intelligent Systems*, 32, 74–80. <https://ieeexplore.ieee.org/document/8267597>.
- Chang, Y.-C., Ku, C.-H., Chen, C.-H. 2019. [Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor](#). *International Journal of Information Management*,



- 48, pp. 263–279, <https://doi.org/10.1016/j.ijinfomgt.2017.11.001>.
- De Matos, C. A., Henrique, J. L., Rossi, C. A. V. 2007. Service recovery paradox: A meta-analysis. *Journal of Service Research*, 10(1), pp. 60–77. <https://journals.sagepub.com/doi/abs/10.1177/1094670507303012>.
- De Pelsmacker, P., van Tilburg, S., Holthof, C. 2018. Digital marketing strategies, online reviews and hotel performance. *International Journal of Hospitality Management*, 72, pp. 47–55, <https://www.sciencedirect.com/science/article/abs/pii/S0278431917305303?via%3Dihub>.
- Devlin, J. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint* arXiv:1810.04805. <https://arxiv.org/abs/1810.04805?amp=1>.
- Bjørkelund, E., Burnett, T. H., Nørvåg, K. 2012, December) A study of opinion mining and visualization of hotel reviews. In *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services* (pp. 229–238). <https://dl.acm.org/doi/abs/10.1145/2428736.2428773>.
- Eriksson, N., Fagerstrøm, A. 2018. The relative impact of wi-fi service on young consumers' hotel booking online. *Journal of Hospitality & Tourism Research*, 42(7), 1152–1169. <https://journals.sagepub.com/doi/abs/10.1177/1096348017696844>.
- Filieri, R., Alguezaui, S., McLeay, F. 2015. Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management*, 51, 174–185, <https://doi.org/10.1016/j.tourman.2015.05.007>.
- Gavilan, D., Avello, M., Martinez-Navarro, G. 2018. The influence of online ratings and reviews on hotel booking consideration. *Tourism Management*, 66, 53–61. <https://doi.org/10.1016/j.tourman.2017.10.018>.
- Gu, B., Ye, Q. 2014. First step in social media: Measuring the influence of online management responses on customer satisfaction. *Production and Operations Management*, 23(4), pp. 570–582, <https://journals.sagepub.com/doi/10.1111/poms.12043>.
- Herrero, Á., San Martín, H., Hernández, J. M. 2015. How online search behavior is influenced by user-generated content on review websites and hotel interactive websites. *International Journal of Contemporary Hospitality Management*, 27, pp. 1573–1597. <https://www.emerald.com/insight/content/doi/10.1108/IJCHM-05-2014-0255/full/html>.
- Hernández-Ortega, B. 2018. Don't believe strangers: Online consumer reviews and the role of social psychological distance. *Information & Management*, 55(1), 31–50. <https://doi.org/10.1016/j.im.2017.03.007>.
- Hu, Y.-H., Chen, Y.-L., Chou, H.-L. 2017. Opinion Mining from Online Hotel Reviews – a Text Summarization Approach. *Information Processing & Management*, 53, pp. 436–449, <https://doi.org/10.1016/j.im.2017.03.007>.
- Huang, S., Peng, W., Li, J., Lee, D. 2013. Sentiment and topic analysis on social media: A multi-task multi-label classification approach. *WebSci '13 Proceedings of the 5th Annual ACM web science conference* (pp. 172–181). <https://dl.acm.org/doi/abs/10.1145/2464464.2464512>.
- Jakob, N., Gurevych, I. 2010. Extracting opinion targets in a single- and cross-domain setting with conditional random fields. *EMNLP '10 Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* (pp. 1035–1045). <https://aclanthology.org/D10-1101.pdf>.
- Lee, Y. J., Xie, K., & Besharat, A. 2016. Management responses to online WOM: Helpful or detrimental?
- King, R. A., Racherla, P., Bush, V.D. 2014. What We Know and Don't Know About Online Word-of-Mouth: A Review and Synthesis of the Literature. *Journal of Interactive Marketing*, 28(3), 167–183. <https://journals.sagepub.com/doi/abs/10.1016/j.intmar.2014.02.001>.
- Korfiatis, N., Stamolampros, P., Kourouthanassis, P., Sagiadinos, V. 2019. Measuring Service Quality from Unstructured Data: A Topic Modeling Application on Airline Passengers' Online Reviews. *Expert Systems with Applications*, 116, 472–486. <https://doi.org/10.1016/j.eswa.2018.09.037>.
- Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint* arXiv:1909.11942. <https://doi.org/10.48550/arXiv.1909.11942>.
- Levy, S. E., Duan, W., Boo, S. 2013. An analysis of one-star online reviews and responses in the Washington, D.C., lodging market. *Cornell Hospitality Quarterly*, 54(1), pp. 49–63, <https://journals.sagepub.com/doi/10.1177/1938965512464513>.
- Liu, B. 2022. *Sentiment analysis and opinion mining*. Springer Nature.

- Liu, K., Xu, L., Zhao, J. 2014. Co-extracting opinion targets and opinion words from online reviews based on the word alignment model. *IEEE Transactions on Knowledge and Data Engineering*, 27(3), 636-650. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6858011>.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*. <https://doi.org/10.48550/arXiv.1907.11692>.
- Lu, W., Stepchenkova, S. 2015. User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software. *Journal of Hospitality Marketing & Management*, 24(2), 119-154. <https://doi.org/10.1080/19368623.2014.907758>.
- Mauri, A.G., Minazzi, R. 2013. Web Reviews Influence on Expectations and Purchasing Intentions of Hotel Potential Customers. *International Journal of Hospitality Management*, 34, pp. 99-107, <https://doi.org/10.1016/j.ijhm.2013.02.012>.
- Melo, A. J. D. V. T., Hernández-Maestro, R. M., Muñoz-Gallego, P. A. 2017. Service quality perceptions, online visibility, and business performance in rural lodging establishments. *Journal of Travel Research*, 56(2), pp. 250-262, <https://journals.sagepub.com/doi/10.1177/0047287516635822>.
- Nie, W. 2000. Waiting: Integrating social and psychological perspectives in operations management. *Omega*, 28(6), 611-629. [https://doi.org/10.1016/S0305-0483\(00\)00019-0](https://doi.org/10.1016/S0305-0483(00)00019-0).
- Noone, B.M., McGuire, K.A. 2013. Pricing in a Social World: The Influence of Non-Price Information on Hotel Choice. *Journal of Revenue & Pricing Management*, 12, 385-401.
- Öğüt, H., Taş, Onur Taş, B. K. 2012. The Influence of Internet Customer Reviews on the Online Sales and Prices in Hotel Industry. *The Service Industries Journal*, 32(2), 197-214. <https://doi.org/10.1080/02642069.2010.529436>.
- Park, S.-Y., Allen, J. P. 2013. Responding to online reviews: Problem solving and engagement in hotels. *Cornell Hospitality Quarterly*, 54(1), 64-73, <https://journals.sagepub.com/doi/10.1177/1938965512463118>.
- Sann, R., Lai, P.-C. 2020. Understanding Homophily of Service Failure Within the Hotel Guest Cycle: Applying Nlp-Aspect-Based Sentiment Analysis to the Hospitality Industry. *International Journal of Hospitality Management*, 91, 102678, <https://doi.org/10.1016/j.ijhm.2020.102678>.
- Schouten, K., Frasincar, F. 2015. Survey on Aspect-Level Sentiment Analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(3), 813-830, <https://ieeexplore.ieee.org/document/7286808>.
- Schuckert, M., Liu, X., Law, R. 2015. Hospitality and Tourism Online Reviews: Recent Trends and Future Directions. *Journal of Travel & Tourism Marketing*, 32(5), pp. 608-621. <https://doi.org/10.1080/10548408.2014.933154>.
- Smith, J. S., Karwan, K. R., Markland, R. E. 2012. An empirical investigation of the effectiveness of an integrated service recovery system. *Operations Management Research*, 5, 25-36. <https://doi.org/10.1007/s12063-012-0063-0>.
- Torres, E. N., Singh, D., Robertson-Ring, A. 2015. Consumer reviews and the creation of booking transaction value: Lessons from the hotel industry. *Journal of Marketing Research*, 55(1), pp. 163-177, <https://journals.sagepub.com/doi/10.1509/jmr.15.0511>.
- Wang, Y., Chaudhry, A. 2018. When and how managers' responses to online reviews affect subsequent reviews. *Journal of Marketing Research*, 55(1), 163-177, <https://journals.sagepub.com/doi/10.1509/jmr.15.0511>.
- Xie, K.L., Zhang, Z., Zhang, Z. 2014. The Business Value of Online Consumer Reviews and Management Response to Hotel Performance. *International Journal of Hospitality Management*, 43, Supplement C, pp. 1-12, <https://www.sciencedirect.com/science/article/abs/pii/S027843191400125X?via%3Dihub>.
- Yu, Y., Li, X., Jai, T.M. 2017. The Impact of Green Experience on Customer Satisfaction: Evidence from Tripadvisor. *International Journal of Contemporary Hospitality Management*, 29(5), 1340-136. <https://doi.org/10.1108/IJCHM-07-2015-0371>.
- Zhao, Y., Xu, X., & Wang, M. 2019. Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management*, 76, 111-121. <https://doi.org/10.1016/j.ijhm.2018.03.017>.