Data Science and Engineering (DSE) Record, Volume 7, issue 1.

Analysis of Travel Application Review Data with Natural Language Processing and Visualization

Rattakarn janya [1] and Arinya Pongwat [2]

[1] Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand
rattakarn_j@cmu.ac.th

[2] College of Arts, Media and Technology, Chiang Mai University, Thailand
arinya.p@cmu.ac.th

Abstract. This independent study aims to analyze user review data and identify key factors influencing user experience (UX) in online travel applications. The study conducts a comparative analysis of three major platforms: Traveloka, Agoda, and Booking.com, using data scraped from the Google Play Store. By leveraging Natural Language Processing (NLP) techniques, the researcher highlights the significance of understanding user needs in the post-pandemic era to enhance digital service efficiency and development. The conceptual framework for categorizing user feedback is based on user experience theories, divided into four primary dimensions: 1) Information Service & Quality, 2) Perceived Benefits, 3) App Performance, and 4) App Design. A dataset comprising 4,035 textual records was collected and subjected to feature extraction for analysis.
Experimental results indicate that Logistic Regression outperformed the other evaluated models, including SVM, Neural Network, Naïve Bayes, Random Forest, and Zero-Shot Learning (ZSL), achieving a classification accuracy of 79.14%. Regarding the thematic analysis, "Information Service & Quality" emerged as the most prominent dimension (26.75%), followed by "Perceived Benefits" (25.46%). Furthermore, in-depth visual analytics using Word Clouds and Co-occurrence Networks revealed that negative reviews were significantly associated with keywords such as "Customer," "Service," and "Refund." These findings suggest that service quality and refund processes are pivotal factors in user decision-making. Consequently, this research serves as a strategic guideline for developers to refine functionalities and better meet the evolving demands of contemporary users.

Keywords: Natural Language Processing, Sentiment Analysis, Travel Application.

1 Introduction

Since 2010, the exponential growth of smartphones has transformed the Information Technology industry, shifting devices from simple communication tools into sophisticated drivers of a Digital Ecosystem [1]. Mobile applications have since become essential daily infrastructure, a trend solidified during the 2015–2019 "New Normal" period, where global downloads reached 120 billion across leading digital economies [2]. This shift is underscored by high engagement rates on platforms like

TikTok, forcing businesses to pivot toward User Experience (UX) to maintain audience engagement [3].

In the travel sector, applications have proliferated to facilitate hotel bookings and travel planning through integrated rating and recommendation systems. While the COVID-19 pandemic necessitated accelerated app development to capture revenue [4], the post-pandemic era saw international travel double in 2022, reaching over 900 million tourists [5]. This surge made travel apps vital tools for price comparison and booking facilitation.

Research indicates that UX and User Interface (UI) are the primary factors in app selection [6]. Notably, 82% of user decisions are "Fast-path," primarily driven by Usefulness (61.2%), while Usability (13.2%) remains the critical variable for user retention [7]. To analyze these behaviors, studies on high-traffic platforms like TripAdvisor and Expedia have increasingly utilized Natural Language Processing (NLP) to refine services based on customer feedback [8][9].

Recognizing this, the researcher aims to apply NLP for Sentiment Analysis and Predictive Modeling. By comparing model accuracies and visualizing user review data, this research seeks to identify the specific application attributes users value most, ultimately providing a roadmap for enhancing future application development and efficiency.

2 Literature Review

2.1 Related Research

Al-Shamaileh and Sutcliffe [7] utilized a mixed-methods approach to analyze mobile app adoption and abandonment. Their findings indicate that 82% of users make "Fast-path decisions," relying on standout features rather than detailed comparisons. Usefulness emerged as the primary driver (61.2% of feedback), as users prioritize real-world problem-solving. While Usability was a secondary factor (13.2%), it remained the leading cause of immediate abandonment if the app was perceived as overly complex or intrusive.

Jiaming Fang et al. [10] applied the S-O-R (Stimulus-Organism-Response) model to investigate travel app engagement through User Acceptance Testing (UAT) with 804 travelers.

- Stimulus: Attributes such as UI attractiveness, security, and ease of use attract users.
- Organism: Positive stimuli trigger perceived Hedonic, Utilitarian, and social benefits, fostering psychological engagement.
- Response: This process results in sustained usage and long-term loyalty.

The study concludes that success requires high "feature quality" to ensure the app is perceived as both useful and satisfying.

Gergana Marinova [11] conducted a performance comparison of algorithms for multi-class classification (Emotion, Finance, Health, etc.) using an English internet dataset. After standardized Data Preprocessing and an 80/20 train-test split, the results showed that Support Vector Machine (SVM) and Neural Network (NN) both achieved the highest accuracy at 0.83. In contrast, the Random Forest (RF) model trailed by 0.18. The study identifies SVM and NN as the most efficient techniques for complex text classification tasks.

2.2 Data Analysis Framework

In this research on Multi-class Classification, the Cross-Industry Standard Process for Data Mining (CRISP-DM) has been applied to ensure that the results are efficient and meet practical functional requirements. The operational process is divided into six phases as follows:

1) **Business Understanding**: This phase involves understanding the context of the data, specifically user reviews, to determine which news categories provide the greatest benefit to the users.

2) **Data Understanding**: This is the process of collecting raw data and verifying its integrity to select the most appropriate dataset for analysis.

3) **Data Preparation**: This phase focuses on Data Cleaning, which is a crucial step for English text. It involves tasks such as removing emojis and links, punctuation removal, and tokenization to prepare the data for Natural Language Processing (NLP).

4) **Modeling**: This involves selecting high-performance algorithms for Multi-class Classification to effectively categorize news articles into predefined groups.

5) **Evaluation**: This phase measures the model's accuracy to ensure system reliability, based on key statistical metrics: Accuracy, Precision, Recall, and F1-Score.

6) **Deployment**: This final stage involves implementing the validated and high-performing model into a real-world environment.

Based on the specialized knowledge of the evolving digital landscape, application marketing has shown significant growth; however, a persistent challenge remains in the misalignment between applications and actual user requirements, which ultimately leads to diminished functional efficiency. This research addresses such issues by collecting and analyzing evaluation data to drive application development, drawing upon the established assessment concepts of Al-Shamaileh and Sutcliffe [7] and Jiaming Fang et al. [10]. Specifically, the Stimulus-Organism-Response (S-O-R) model [10] serves as the foundation for the research conceptual framework, focusing on users of travel applications with similar functionalities, namely Traveloka, Agoda, and Booking.com. Data for this analysis is sourced from the Google Play Store and categorized into three core components:

- Stimulus, which encompasses App Design Attributes (UX/UI), App Performance, and Information Service & Quality.
- Organism, representing the user's internal perception through Perceived Benefits.
- Response, manifested as User Feedback and Reviews regarding their experiences. These interrelationships are further illustrated in Figure 1 . Research Conceptual Framework.

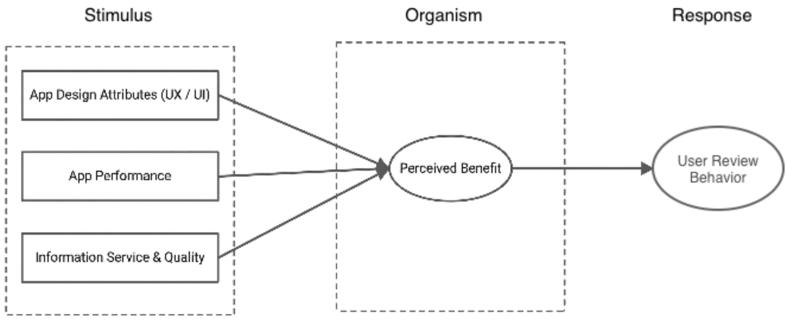

Figure 1. Research Conceptual Framework

The analytical process begins by converting textual review data into a quantitative format using the TF-IDF (Term Frequency-Inverse Document Frequency) technique to determine the statistical weight and significance of each term. Subsequently, the study conducts a performance comparison across various algorithms within the Supervised Learning category, including Support Vector Machine (SVM), Neural Network (NN), Random Forest Classifier [11], Naïve Bayes (NB), and Logistic Regression. In addition, Zero-Shot Learning (ZSL) is incorporated to evaluate its capability in classifying reviews without extensive task-specific training. For the experimental setup, the dataset is partitioned into two distinct parts: a Training Set (80%) for model development and a Test Set (20%), serving as unseen data to evaluate the model's predictive capabilities in real-world scenarios. Finally, the prediction results from each algorithm are validated against standard statistical metrics, specifically Precision, Recall, and F1-Score, to determine the overall efficiency and accuracy for each category. This comprehensive analysis aims to classify travel application reviews effectively, ensuring the outcomes strictly align with the defined research objectives.

3 Data and Methodology

3.1 Selection Criteria for Application Analysis

In gathering the dataset for user opinion analysis, the researcher established specific criteria to select three prominent travel applications: Traveloka, Agoda, and Booking.com. These applications were chosen based on their extensive user bases and high popularity rankings on the Google Play Store. As market leaders in the accommodation and travel booking industry, these three platforms provide a sufficient volume of review data required for effective model training. Furthermore, the selected applications share consistent functional features, such as property search engines, flight booking systems, and online reservation processes, ensuring a standardized baseline for comparative analysis.

3.2 Selection and Data Collection Approach for User Reviews

A total of 4,035 English-language user reviews were extracted from the Google Play Store for Traveloka, Agoda, and Booking.com using a Web Scraper extension. The data collection focused on the 2022–2023 period to reflect post-pandemic user behavior. Strict screening criteria were applied to ensure data quality: (1) filtering out non-substantive single-word comments (e.g., "good," "bad"), (2) excluding reviews unrelated to app performance (e.g., specific hotel services), and (3) performing comprehensive Data Cleansing to remove ambiguous or repetitive text, ensuring the dataset was optimized for NLP analysis.

3.3 Categorization and Classification

The unstructured reviews were systematically categorized into four dimensions based on the S-O-R framework. To minimize bias, classification was guided by specific keyword detection:

- App Design Attributes (UX/UI): Keywords such as "UI," "Interface," "Design," and "Layout."
- Performance: Technical terms like "Error," "Search," "Login," and "Loading."
- Information Service & Quality: Transactional terms including "Service," "Customer," "Refund," and "Booking."
- Perceived Benefit: Value-oriented words such as "Price," "Experience," "Discount," and "Worth."

The final dataset demonstrated a balanced distribution across the four categories: Information Service & Quality (26.75%), Perceived Benefit (25.46%), App Performance (24.66%), and App Design Attributes (23.13%). This consistent distribution (max variance 3.62%) provides a robust foundation for model training, as illustrated in Table 1.

Table 1. Distribution of user review counts across categorized dimensions

Category	Counts
Information Service & Quality	1,079
Perceived Benefit	1,127
App Performance	995
App Design Attributes	933

Statistically, this distribution represents a balanced dataset, eliminating the need for further data balancing techniques such as Oversampling (e.g., SMOTE) or Under sampling. Since each class contains a sufficient and equitable volume of data, the model can impartially learn the distinctive features of each category without bias. Consequently, this ensures the reliability of the Accuracy and F1-score metrics, ultimately enhancing the model's performance in accurately classifying user sentiments within travel applications.

3.4 Data Preprocessing

To prepare raw text for machine learning, a systematic preprocessing pipeline was implemented: (1) Cleaning, including special character/emoji removal and contraction expansion; (2) Standardization, via lowercasing and filtering to alphabetic characters (a-z); (3) Tokenization, to segment text into word units; and (4) Stop Words Removal, to eliminate non-semantic terms. Finally, the processed text was converted into a numerical feature matrix using TF-IDF (Term Frequency-Inverse Document Frequency) for model training.

3.5 Data Analysis and Modeling

This study evaluates the performance of multiple Supervised Learning models, including SVM, Neural Network (NN), Naïve Bayes (NB), Logistic Regression, and Random Forest, in comparison with a Zero-Shot Learning (ZSL) approach. By integrating these algorithms with Natural Language Processing (NLP), the research aims to accurately classify feedback into four primary dimensions: App Design (UX/UI), Performance, Information & Service Quality, and Perceived Benefit. This comparative analysis seeks to identify the optimal model for transforming unstructured reviews into actionable insights, thereby providing a reliable foundation for enhancing travel applications.

4 Results

4.1 Classification Results

After developing the classification models using TF-IDF for feature extraction, the performance of each algorithm was evaluated. The comparative results, including precision, recall, and F1-score for each model, are summarized in Table 2 and Table 3

Table 2. Performance Comparison of Classification Models.

Model Name	Accuracy	Precision(avg)	Recall(avg)	F1-score
Logistic Regression	0.791	0.79	0.79	0.79
Naive Bayes	0.770	0.77	0.77	0.77
SVM	0.789	0.79	0.78	0.78
Neural Network	0.728	0.73	0.73	0.73
Random Forest	0.758	0.77	0.76	0.76
Zero-Shot Learning (ZSL)	0.477	0.53	0.32	0.38

Table 3. Performance Comparison of Classification Models.

Model	Category	Precision	Recall	F1-Score
TF-IDF + Logistic Regression	App Design Attributes	0.82	0.74	0.78
	App Performance/Capability	0.75	0.73	0.74
	Information Service & Quality	0.86	0.85	0.86
	Perceived Benefits	0.73	0.82	0.76
TF-IDF + Naive Bayes	App Design Attributes	0.82	0.72	0.77
	App Performance/Capability	0.76	0.65	0.70
	Information Service & Quality	0.78	0.88	0.83
	Perceived Benefits	0.72	0.80	0.76
TF-IDF + SVM	App Design Attributes	0.82	0.74	0.78
	App Performance/Capability	0.73	0.74	0.73
	Information Service & Quality	0.85	0.85	0.85
	Perceived Benefits	0.73	0.79	0.76
TF-IDF + Random Forest	App Design Attributes	0.83	0.71	0.77
	App Performance/Capability	0.72	0.65	0.68
	Information Service & Quality	0.85	0.82	0.83
	Perceived Benefits	0.65	0.85	0.74
TF-IDF + Neural Network	App Design Attributes	0.73	0.69	0.71
	App Performance/Capability	0.65	0.64	0.65
	Information Service & Quality	0.86	0.81	0.83
	Perceived Benefits	0.66	0.76	0.71
ZSL	App Design Attributes	0.80	0.14	0.24
	App Performance/Capability	0.35	0.67	0.46
	Information Service & Quality	0.64	0.09	0.16
	Perceived Benefits	0.35	0.58	0.43

Experimental Design: User reviews were classified using TF-IDF features across multiple models: NB, SVM, NN, Logistic Regression, Random Forest, and ZSL. Performance was assessed via Accuracy, Precision, Recall, and F1-score, employing Macro-averaging to ensure balanced evaluation across all categories.

According to Table 2. Performance Comparison of Classification Models and Table 3 (Accuracy Comparison for User Review Classification), the Logistic Regression model demonstrated the highest stability and performance with an accuracy of 0.791. Conversely, the Zero-Shot Learning (ZSL) model yielded the lowest performance, with a significantly lower accuracy of 0.477.

4.2 Word Cloud Visualization

Following the Data Cleansing and Word Tokenization of the 4,035-message dataset, the results were analyzed and visualized using Word Clouds to represent term frequency and core themes, as detailed in Figure 2.

Figure 2. Word Cloud of Frequent User Feedback Terms

According to Figure 2, the most prominent and frequently occurring terms reflect the primary concerns of the users. The term “app” emerges as a central theme, indicating that feedback is directly focused on the platform itself. The keyword “booking” signifies that the "reservation process" is the core activity discussed, highlighting it as a critical function within this application group. Furthermore, the high prevalence of “customer” and “service” demonstrates that users place significant importance on "customer support" and "service quality."

To provide a more granular analysis, the researcher categorized the Word Clouds by dimension, as shown in Figure 3, 4, 5 and 6.

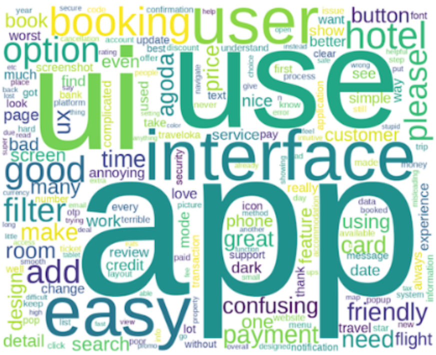

Figure 3. Word Cloud of App Design

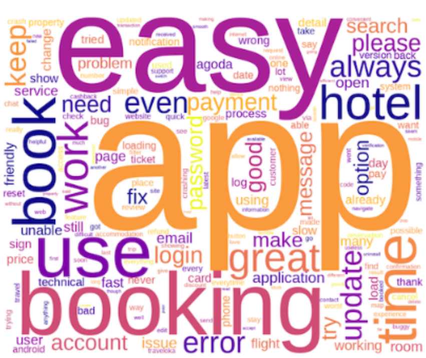

Figure 4. Word Cloud of Performance

Figure 5. Word Cloud of Service & Quality

Figure 6. Word Cloud of Benefits

Based on the Word Clouds Figure 3, 4, 5, and 6. the synthesized results are summarized in the table below

Table 4. Key terms categorized by dimension.

Category	Word
Information Service & Quality	service, customer, booking, refund, money, help, support, call, email, agent, reservation, ticket, payment, charge, transaction, confirm, information, staff, policy
App Design Attributes	UI, interface, design, navigation, button, mode, user, friendly, layout, look, font, color, icon, display, screen, menu, option, tab
App Performance	load, slow, crashing, search, work, open, error, glitch, bug, update, connection, quickly, fast, stable, freeze, battery, storage
Perceived Benefits	experience, useful, benefit, price, great, easy, helpful, value, discount, cheap, deals, travel, worth, efficient, simple, nice, best, amazing

Based on the importance weights presented in Table 4, the key factors driving model classification can be summarized as follows. Regarding Information Service & Quality, terms such as "service," "customer," and "refund" indicate that consumers prioritize staff responsiveness and transactional accountability, while "money" reflects high sensitivity toward financial data accuracy. For App Design Attributes, keywords like "UI," "interface," and "friendly" show that visual appeal and accessibility directly impact initial satisfaction, with "dark mode" highlighting a demand for modern usability trends. In terms of App Performance, negative influence stems from terms like "slow," "load," and "crashing," identifying system stability and search speed as critical technical factors. Finally, Perceived Benefits are driven by "useful," "price," and "worth," reflecting a primary focus on value and discounts, while "experience" suggests that users prioritize overall travel convenience alongside competitive pricing.

4.3 Co-occurrence Network

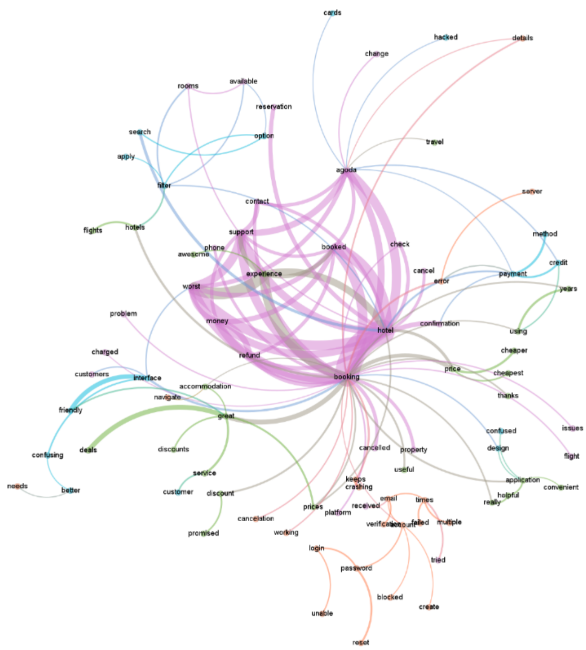

Figure 7. Results of Word Co-occurrence Network Analysis

This network visualization demonstrates that user feedback is highly interconnected through co-occurrence, maintaining a balanced distribution across all dimensions. The Information Service & Quality (Pink) and Perceived Benefits (Green) clusters form the core of the network with the densest connections, suggesting that superior service quality directly drives the perception of user benefits. Specifically, Information Service & Quality frequently overlaps with other categories via terms like "service," "customer," and "booking," indicating that service quality is a primary decision factor often discussed alongside technical or design aspects. Perceived Benefits shows a strong affinity with Service and Design (Blue), where terms like "nice experience" and "good price" typically co-occur with user-friendly interface mentions. In contrast, App Performance (Orange) elements such as "server," "load," and "crashing" tend to cluster more distinctly; this isolation suggests that users experiencing technical failures focus their complaints specifically on system stability rather than aesthetics or service. Finally, the clear linkage between App Design Attributes (e.g., "UI," "interface") and the Benefits cluster confirms that user-friendly design significantly enhances the user's perceived value of the application.

5 Conclusion

This study analyzes user reviews to identify key factors for optimizing travel application development and meeting user needs. This chapter summarizes the research findings, discusses the study's limitations, and addresses the challenges encountered during the research process.

5.1 Discussion

The primary objective of this study was to analyze key factors influencing user experiences within travel applications. By classifying user sentiments, the research aims to provide actionable insights for developing highly efficient business strategies. The researcher collected 4,035 user reviews via web scraping from the Google Play Store, focusing on the most popular applications: Traveloka, Agoda, and Booking.com, covering the period from 2022 to 2023.

Performance testing revealed that the Logistic Regression model achieved an overall accuracy of 79.18%. When examining categorical performance via the radar chart in Figure 8, a balanced classification across all dimensions is observed.

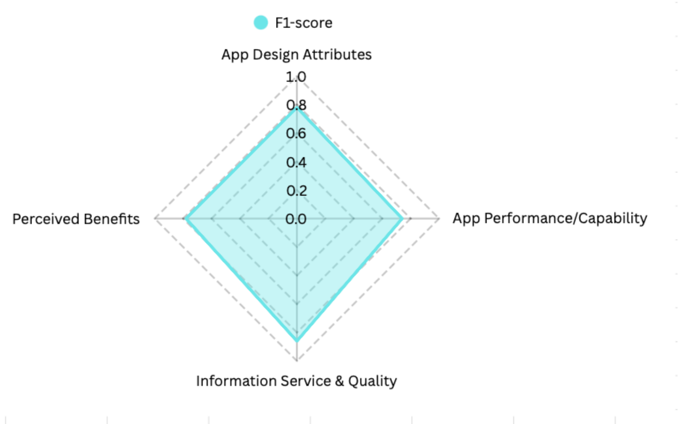

Figure 8. Radar Chart of Logistic Regression Model by Category

Furthermore, the radar chart analysis in Figure 8 reveals that the Logistic Regression model, the best-performing model in this study, demonstrates comprehensive and balanced classification capabilities across all categories, with F1-scores ranging from 0.74 to 0.86. A more detailed breakdown of these results is illustrated in the Confusion Matrix shown in Figure 9.

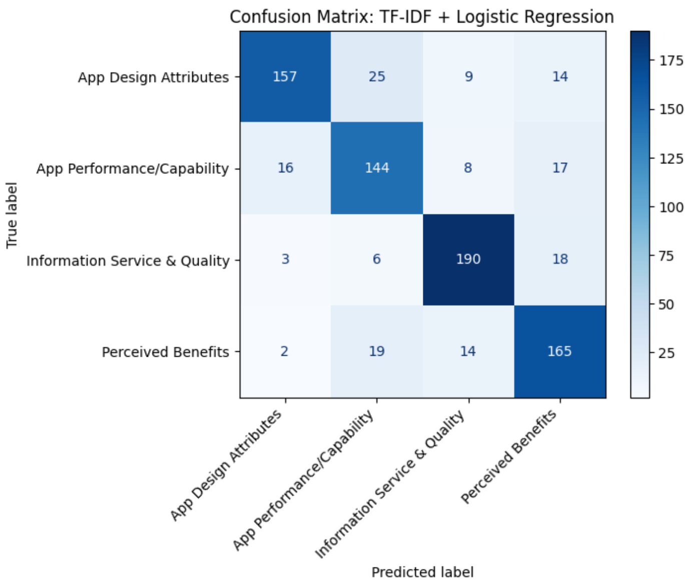

Figure 9. Confusion Matrix of the Logistic Regression model by category

The analysis shows that the most accurately predicted category is Information Service & Quality, which achieved 190 True Positives and the highest F1-score of 0.86. This high performance is attributed to the presence of distinct and specific keywords such as 'Service', 'Refund', and 'customer'. This was followed by Perceived Benefits and App Design Attributes, with 165 and 157 correct predictions and F1-scores of 0.76 and 0.78, respectively. The model effectively classified contexts related to positive user experiences and general usability issues, such as 'User friendly UI'and 'very nice experience'.

In contrast, App Performance/Capability recorded the lowest F1-score of 0.74, with 144 correct predictions. The lower performance in this category stems from semantic overlap in user feedback. Specifically, App Design Attributes showed the highest overlap with the performance category (25 instances), as users often associate UI and accessibility with system functionality. Furthermore, 19 instances of Perceived Benefits overlapped with performance, likely because error-free and rapid application speed lead users to perceive the app as highly beneficial. Additionally, 18 instances of Information Service & Quality were misclassified as Perceived Benefits, reflecting that efficient service directly influences the overall perceived value of the travel application. The resulting sentiment classification is illustrated in Figure 10, with further analysis detailed below.

Review Content	Predicted Category
quickly search.	App Performance/Capability
good service	Information Service & Quality
experience is not good	Perceived Benefits
cant find the cart button	App Design Attributes
very bad customer service, not help at all	Information Service & Quality
User friendly UI	App Design Attributes
App won't load	App Performance/Capability
very nice experience thankyou	Perceived Benefits

Figure 10. Classification Accuracy by Category (F1-score)

- **Information Service & Quality (F1-score: 0.86)**: This category achieved the highest accuracy due to the presence of highly specific keywords such as "service," "customer," "booking," and "refund." These terms carry significant TF-IDF weighting, which reduces semantic ambiguity. As illustrated in the Word Cloud (Figure 5) , the model effectively distinguished both positive and negative sentiments (e.g., "good service" vs. "bad customer service") with minimal error. These insights are vital for developers to automate and accelerate customer support screening.
- **App Design & Perceived Benefits (F1-score: 0.78)**: The model demonstrated strong capability in identifying design-related feedback (e.g., "User friendly UI") and user-perceived value (e.g., "very nice experience").
- **App Performance (F1-score: 0.74)**: Despite being the lowest score, the model remained reliable in detecting technical issues and capabilities, such as search speed and loading failures.

The experimental results indicate that Logistic Regression, when paired with TF-IDF vectorization, is the most effective approach for this dataset. Its superior performance suggests that the linear relationship within this specific text distribution is better captured by Logistic Regression than by other models. In conclusion, this model provides a robust foundation for transforming unstructured user feedback into actionable insights to enhance travel application quality and user satisfaction.

5.2 Theoretical Foundation

1) Identification of UX Dimensions and Post-Pandemic Trends

The study effectively categorizes the primary factors influencing user experience into four dimensions: Information Service & Quality (26.75%), Perceived Benefits (25.46%), App Performance (24.66%), and App Design (23.13%). These results indicate that in the post-pandemic era, users prioritize information and service quality above all other factors (Table 1). Visual analytics via Word Clouds and keyword frequency tables (Table 4) further illustrate this, as the "Service" dimension is strongly reflected through terms such as "Service" and "Refund," while "Perceived Benefits" are associated with "Useful" and "Price." These findings satisfy the research objective of identifying concrete user concerns from raw textual data. Additionally, the Co-occurrence Network (Figure 7) demonstrates that App Design and Service Quality are directly linked to the user's perception of benefits, providing insight into the underlying mechanisms of how these factors interact within the travel application ecosystem.

2) Theoretical Contributions and Alignment with Prior Research

The findings align with the research by Jiaming Fang et al. [10], which posits that feature performance and quality are primary drivers of user satisfaction. The experimental results from the Logistic Regression model, which achieved an overall accuracy of 79.18%, confirm that the "Information Service & Quality" category yielded the highest F1-score of 0.86. This suggests that users express their opinions on service quality with such clarity that the model can learn and classify this dimension more accurately than others. This outcome emphasizes that users value functional utility and service reliability over mere aesthetic design (which yielded a lower F1-score of 0.78). From a Theoretical Foundation perspective, this study validates the Stimulus-Organism-Response (S-O-R) framework; specifically, Information and Service Quality serve as the critical Stimulus that directly impacts the Organism (the user’s internal perception of utility and satisfaction), ultimately leading to the Response (textual feedback and long-term application engagement).

3) Comparative Performance of Classification Technologies

In terms of algorithmic efficiency, the results show that Logistic Regression achieved the highest accuracy (0.791), demonstrating greater stability compared to more complex models like Neural Networks (0.728). This finding contrasts with the research by Gergana Marinova [11], which reported that Neural Networks could achieve an efficiency of 0.83 on different datasets. This discrepancy can be attributed to the size of the current dataset (4,035 records), which may not be sufficient for Deep Learning models to reach their full potential. Conversely, linear models combined with TF-IDF feature extraction prove to be more suitable and effective for this specific scale of textual data in the travel application domain.

5.3 Limitations of this Study

1) Contextual and Industry Limitation

The findings of this research are specifically derived from the travel application context. Since user sentiments and linguistic patterns vary significantly across different industries, the current dataset and feature extraction techniques may not be directly applicable to other application categories without further adaptation

2) Linguistic and Cultural Limitation

A primary limitation is the exclusive focus on English-language reviews. This scope may not fully capture regional cultural contexts or the unique pain points of local users, such as Thai speakers, whose specific grievances might not be fully represented in international datasets.

3) Granularity and Technical Sophistication

While the current model effectively classifies broad dimensions, it lacks the granularity to isolate specific urgent sub-categories, such as "Refund" or "Payment Issues." Furthermore, as this study utilizes traditional classification methods, it may not capture the deep, nuanced semantics that more advanced architectures like BERT or Transformer-based models could provide.

5.4 Recommendations

Based on the findings of this study, the researcher provides the following recommendations:

1) Strategic and Practical Recommendations

Regarding the strategic implications of this study, it is important to note that the current findings are specifically derived from the travel application context. Since user sentiments and linguistic patterns vary significantly across different industries, future implementations should adapt both the dataset and feature extraction techniques to the unique characteristics of the target application category to ensure sustained classification accuracy. Furthermore, a primary limitation of this research was its exclusive focus on English-language reviews. Future studies should incorporate local languages, such as Thai, to more effectively capture regional cultural contexts and identify the unique pain points of local users, which may not be fully represented in international datasets.

2) Recommendations for Future Research

While the current model effectively classifies broad dimensions, future research should apply advanced NLP techniques to achieve a more granular classification of urgent sub-categories. For instance, specifically isolating "Refund" or "Payment Issues" from general feedback would enable developers to prioritize critical grievances and implement immediate solutions. To achieve this, future experiments should compare performance with more sophisticated architectures, such as Deep Learning or Transformer-based models like BERT. These advanced models can provide deeper insights into complex user sentiments and nuanced semantics, allowing for a more precise identification of critical issues. Such technological advancements will be vital for sustaining long-term user engagement and fostering enduring loyalty in an increasingly competitive market.

References

1. Prakobtham, B. (2010). Smartphone business trends in 2010. Executive Journal, Bangkok University, 30(2), 197-202.
2. Ms.Nokyung. (2020, January 23). Statistics on "App" usage reflect the New Normal since 2019,with 120 billion downloads worldwide. Marketing Oops! https://www.marketingoops.com/digital-life/report-app-usage/
3. Wynnsoft Solution. (2018, January 23). Modern marketing and Mobile Applications.
4. Yawised, K., Apasrawirote, D., Chatrangsan, M., & Muneesawang, P. (2021). Factors affecting SMEs' intention to adopt a mobile travel application based on the Unified Theory of Acceptance and Use of Technology (UTAUT-2). Emerging Science Journal, 5(Special Issue), 195-212. https://doi.org/10.28991/esj-2021-SP1-014.
5. Majumdar, R. (2023). International travel surged in 2022, UN says. DW. https://www.dw.com/en/international-travel-surged-in-2022-un-says/a-64430320
6. Putra, K. C., Dewi, I. C., & Budi, A. S. (2022). Usability factors that drive continued intention to use and loyalty of mobile travel application. Heliyon, 8(9), e10620, https://doi.org/10.1016/j.heliyon.2022.e10620.
7. Al-Shamaileh, Y., & Sutcliffe, A. (2023). Fast-path decision making: Why users adopt and abandon mobile apps. International Journal of Human-Computer Studies, 170, 102964. https://doi.org/10.1016/j.ijhcs.2022.102964
8. Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. Tourism Management, 58, 51-65, https://doi.org/10.1016/j.tourman.2016.10.001.
9. Chang, Ku, & Chen. (2020). Using deep learning and visual analytics to explore hotel reviews and responses. Tourism Management, 80,
10. Fang, J., Zhao, Z., Wen, C., & Wang, R. (2017). Design and performance attributes driving mobile travel application engagement. Information & Management, 54(6), 796-808. https://doi.org/10.1016/j.ijinfomgt.2017.03.003
11. Marinova, G. (2024). Comparison of algorithms for text classification. AIP Conference Proceedings, 3105(1), 060002, https://doi.org/10.1063/5.0210081.