International Journal on Computational Engineering

Journal Homepage www.comien.org/index.php/comien

Predicting Customer Loyalty in the Airline Industry: A Machine Learning Approach Integrating Sentiment Analysis and User Experience

Rumana Shahid [a], Md Abu Sufian Mozumder [b,*], Md Murshid Reja Sweet [a], Mehedi Hasan [c], Mahfuz Alam [d], Mohammad Anisur Rahman [e], Mani Prabha [d], Md Arif [a], Md Parvez Ahmed [f], Md Rafiqul Islam [d]

[a] *Department of Management Science and Quantitative Methods, Gannon University, USA*
[b] *College of Business, Westcliff University, Irvine, California, USA*
[c] *Master of Science, Management- Business Analytics, St. Francis College, USA*
[d] *Department of Business Administration, International American University, Los Angeles, California, USA*
[e] *Department of Marketing & Business Analytics, Texas A&M University-Commerce, USA*
[f] *Master of Science in Information Technology, Washington University of Science and Technology, USA*
Corresponding author: [*]*m.mozumder.158@westcliff.edu*

***Abstract*— Ensuring customer loyalty is crucial for the success of any airline service provider in today's competitive environment. This study employs machine learning techniques to predict the likelihood of customers revisiting airline services, emphasizing the role of emotional connections in fostering loyalty. By analyzing feedback comments and satisfaction ratings, we explore how sentiments expressed by customers correlate with their propensity to return. Using sentiment analysis and features extracted through the Linguistic Inquiry and Word Count (LIWC) methodology, we categorize sentiments into various dimensions, integrating these with user experience (UX) elements for a comprehensive predictive model. Our methodology includes a robust data collection process, involving an initial survey of 17,000 valid responses and a follow-up survey one year later. We evaluate multiple classifiers, including Decision Tree, Random Forest, and XGBoost, through five-fold cross-validation. Results reveal that XGBoost achieves the highest accuracy of 85% in predicting return visits, highlighting the predictive power of machine learning in understanding customer behavior. These findings offer significant insights for airlines, enabling them to tailor services and strategies to enhance customer satisfaction and loyalty. Our study underscores the importance of sentiment analysis and UX in predicting customer loyalty, providing a roadmap for future research and practical applications in the airline industry.**

***Keywords*— Customer loyalty; airline industry; machine learning; sentiment analysis; customer satisfaction; user experience (UX); return visits; predictive modeling.**

Manuscript received 10 Jan. 2024; revised 14 Mar. 2024; accepted 18 Apr. 2024. Date of publication 30 Jun. 2024.

I. Introduction

In today's fiercely competitive landscape, ensuring customer loyalty is paramount for the success and sustainability of any service provider, particularly within the dynamic and ever-evolving domain of the airline industry. Understanding the intricate factors that influence customers' intentions to revisit services is not only pivotal but also serves as a catalyst for continual improvement and innovation. In response to this imperative, our study embarks on a quest to estimate the likelihood of customers' return visits to airline services through the lens of machine learning.

Customer loyalty is not merely a function of transactional interactions; rather, it hinges significantly on the emotional connections forged between customers and service providers. Emotions play a pivotal role in shaping attitudes, behaviors, and intentions, exerting a profound influence on customers' decisions to revisit services. Recognizing this fundamental truth, our study delves deep into the realm of affective expressions as encapsulated within customer feedback comments. Drawing on established theories of affective processes and user experience (UX), we endeavor to decipher how sentiments articulated by customers correlate with their propensity to return to airline services.

Central to our investigation are the features extracted from feedback comments and satisfaction ratings garnered from previous service interactions. These features serve as the bedrock upon which our predictive models are built, offering invaluable insights into the underlying factors driving return visit behavior. Leveraging sentiment analysis techniques, we categorize sentiments into distinct dimensions, ranging from overall satisfaction to nuanced affective expressions, thus unraveling the multifaceted tapestry of customer experiences.

The methodology underpinning our study is both rigorous and comprehensive. Through a meticulously orchestrated online survey initiative, we engage with a vast cohort of customers who have recently availed themselves of airline services. By meticulously processing and validating responses, we curate a robust dataset that forms the backbone of our analysis. Moreover, a follow-up survey conducted one year later provides invaluable insights into actual revisit data, thereby offering a nuanced understanding of customer behavior over time.

Within our analysis, we employ a myriad of machine learning classifiers, each meticulously evaluated through a rigorous five-fold cross-validation procedure. From Decision Tree to XGBoost, these classifiers undergo a comprehensive assessment to determine their efficacy in predicting customers' return visits. The results gleaned from this analysis offer tantalizing glimpses into the predictive power of machine learning in discerning patterns within customer feedback.

As we navigate through the labyrinth of data and analysis, our study unravels profound implications for the airline industry. By shedding light on the nuanced interplay between customer sentiments, satisfaction, and return visit behavior, we empower airlines to make informed decisions and strategic interventions that resonate with their customers. From tailoring service offerings to amplifying customer engagement strategies, our findings offer a roadmap for airlines seeking to foster long-term loyalty and sustainable growth in an increasingly competitive marketplace.

Ensuring the return of customers is a cornerstone of success for any service provider, particularly in the competitive landscape of the airline industry. Understanding the factors influencing customers' intentions to revisit services is pivotal, prompting extensive investigation into both motivating and inhibiting factors. This study endeavors to estimate the likelihood of customers' return visits to airline services using a machine learning approach. By harnessing feedback comments and satisfaction ratings garnered from previous service interactions, we aim to predict return visit behavior with an accuracy of 85%.

Customer loyalty hinges significantly on their propensity for return visits, with emotions playing a pivotal role in shaping attitudes and intentions. Recognizing this, our study delves into users' affective expressions as captured in feedback comments. Drawing on established theories of affective processes and user experience (UX), we explore how sentiments expressed by customers correlate with their likelihood of returning to airline services. Features extracted from feedback comments and satisfaction ratings serve as key predictors of return visit behavior. Utilizing sentiment analysis techniques, we categorize sentiments into distinct dimensions, including total, basic, UX, and a combined basic and UX dimension [1]. These dimensions encapsulate various aspects of user experiences and sentiments, ranging from overall satisfaction to nuanced affective expressions.

These findings hold profound implications for the airline industry, shedding light on the nuanced interplay between customer sentiments, satisfaction, and return visit behavior. By leveraging machine learning techniques, airlines can discern patterns within feedback comments to anticipate customers' future actions more accurately. Moreover, the emphasis on user experience underscores the importance of catering to both pragmatic and hedonic aspects of service delivery to foster customer loyalty.

Customer loyalty has long been recognized as a crucial determinant of success for service providers, particularly within the highly competitive airline industry. According to Reichheld and Sasser [2] retaining customers is more cost-effective than acquiring new ones, making loyalty a critical focus for businesses. This principle holds true in the airline sector, where customer satisfaction and repeat patronage directly influence profitability and market share (Han, Back, & Barrett [8].

Emotional connections between customers and service providers play a pivotal role in fostering loyalty. Studies have shown that emotions significantly impact customers' attitudes, behaviors, and intentions [1]. In the context of airlines, the emotional quality of customer service experiences can deeply affect satisfaction and loyalty [5]. This emotional engagement can be effectively measured through sentiment analysis of customer feedback, providing insights into the affective dimensions of customer experiences [13].

Sentiment analysis, the process of extracting and analyzing subjective information from text data, has emerged as a powerful tool in understanding customer sentiments and predicting behavior [15]. In the airline industry, sentiment analysis of customer feedback comments can uncover nuanced emotional responses that influence loyalty. The use of tools like the Linguistic Inquiry and Word Count (LIWC) methodology allows for a detailed examination of emotional expressions in text, facilitating a deeper understanding of customer satisfaction and intentions [16].

The application of machine learning in predictive modeling has revolutionized how businesses forecast customer behavior. Machine learning algorithms, such as XGBoost and Random Forest, are particularly effective in handling large datasets and identifying patterns that may not be apparent through traditional statistical methods [4]. These algorithms have shown significant promise in predicting customer loyalty by analyzing features derived from feedback comments and satisfaction ratings [9].

User Experience (UX) has been identified as a critical factor influencing customer satisfaction and loyalty. According to [10] UX encompasses both pragmatic and hedonic aspects of service interactions, which collectively shape overall satisfaction. In the airline industry, positive UX elements, such as ease of booking and in-flight comfort, contribute significantly to customer retention [11].

Integrating sentiment analysis with UX elements provides a holistic approach to predicting customer loyalty. By examining both emotional expressions and user experience factors, predictive models can achieve higher accuracy in forecasting repeat patronage. This integrated approach has been validated in various studies, highlighting its effectiveness in understanding and predicting customer behavior [14].

The methodological rigor in data collection and validation is paramount for the accuracy of predictive models. The use of professional survey agencies and rigorous follow-up procedures, as described in this study, ensures the reliability of the dataset [7]. Moreover, employing cross-validation techniques in evaluating machine learning classifiers enhances the robustness of the findings [12].

The findings of this study offer valuable insights for the airline industry, emphasizing the importance of emotional connections and user experience in fostering customer loyalty. However, the study also highlights the need for further research incorporating additional data sources, longitudinal analyses, and advanced machine learning techniques [6]. By addressing these areas, future research can provide a more comprehensive understanding of the complex dynamics influencing customer loyalty in the airline industry.

II. Material and Methods

In November 2017, a professional survey agency embarked on an online survey initiative to gauge customer satisfaction with airline services. The survey was distributed through a pop-up window to 30,000 customers who had booked their flights online and had utilized airline services within the previous six months. These customers were instructed to leave their feedback comments and satisfaction ratings within two to three weeks of their last flight. The survey agency meticulously processed the responses, eliminating invalid ones as well as those from flights that were delayed or cancelled. This rigorous validation process resulted in 17,000 valid responses, representing 57.9% of the total distributed surveys. Fast forward one year to November 2018, the survey agency followed up with the customers who had provided these validated responses to collect actual revisit data. These customers were queried on whether they had used airline services in the past year. From the 17,000 initial respondents, 13,000 confirmed they had indeed used airline services again. Further, these 13,000 respondents were asked if they had chosen the same airline service as indicated in the initial survey, thereby providing vital follow-up data on customer loyalty and the propensity for repeat usage of the airline services. This follow-up not only shed light on the effectiveness of the airlines' customer retention strategies but also offered a deeper understanding of customer behavior over a longer period.

A. Dataset Features

In the endeavor to predict customers' return visits to airline services, this study employed two primary features: satisfaction ratings and elements extracted from feedback comments, as delineated in Table 1. A crucial aspect of this analysis involved conducting sentiment analysis on the comments to delve into users' emotional expressions. To achieve this, the Linguistic Inquiry and Word Count (LIWC) methodology was utilized to scrutinize the sentiment embedded within the comments. The LIWC facilitated the extraction of sentimental expressions, which were subsequently employed as features in the study's analysis. Two main approaches were adopted for feature selection: the Total Dimension and the Basic Dimension.

Under the Total Dimension approach, all extracted sentimental factors were utilized as features, offering a comprehensive overview of sentiment across the comments. Meanwhile, the Basic Dimension approach combined descriptive and affective factors. Descriptive factors encompassed summary and linguistic variables, including metrics such as 'Clout' and 'Authentic' derived from LIWC. Alongside these, affective factors were integrated, comprising positive and negative emotion dimensions as well as the 'negate' variable, enabling a nuanced exploration of sentiment within the comments. Moreover, the study incorporated User Experience (UX) elements, as delineated , into the analysis under the UX Dimension. These elements provided additional insights into factors influencing customer return visits beyond mere sentiment analysis. Lastly, a hybrid approach merging the Basic and UX dimensions was examined to discern any synergistic effects between sentiment and user experience on customer revisit intentions. By leveraging these multifaceted dimensions, the study aimed to unravel the intricate interplay between customer satisfaction, sentiment, and user experience in shaping repeat patronage of airline services.

TABLE I
PRESENTS THE CATEGORIES OF EMOTIONAL EXPRESSIONS, INCLUDING SATISFACTION RATINGS AND FEATURES EXTRACTED USING LIWC

Sentimental expressions and construct	**Total Dimension**	**Basic Dimension**	**UX Dimension**	**Basic and UX Dimension**
Satisfaction	O	O	O	O
Word count	O			
Authentic	O	O	O	O
Tone	O	O		O
affect	O	O		O
anger	O			
compare	O		O	O
work	O		O	O
leisure	O		O	O
affiliation	O			
reward	O			

Using various machine learning classifiers commonly applied in binary classification tasks, we aimed to determine whether a given user who initially utilized an airline service also revisited the same service. Seven classifiers, namely Decision Tree, Gaussian Naive Bayes, Logistic Regression, Random Forest, KNeighbors, Support Vector Machine, and XGBoost, underwent evaluation through a five-fold cross-validation procedure. In this process, the dataset was partitioned into five folds, with one-fold serving as the test set in each iteration. The performance of the classifiers was assessed based on the average accuracy across the five iterations, serving as the final evaluation metric.

III. Result and Discussion

In our study, we delve into the crucial aspect of customer return visits and employ a machine learning methodology to predict these visits. Our research builds upon previous findings in service sciences, particularly emphasizing the significance of user experience (UX) as a determinant of customer satisfaction and intention to revisit. We integrate

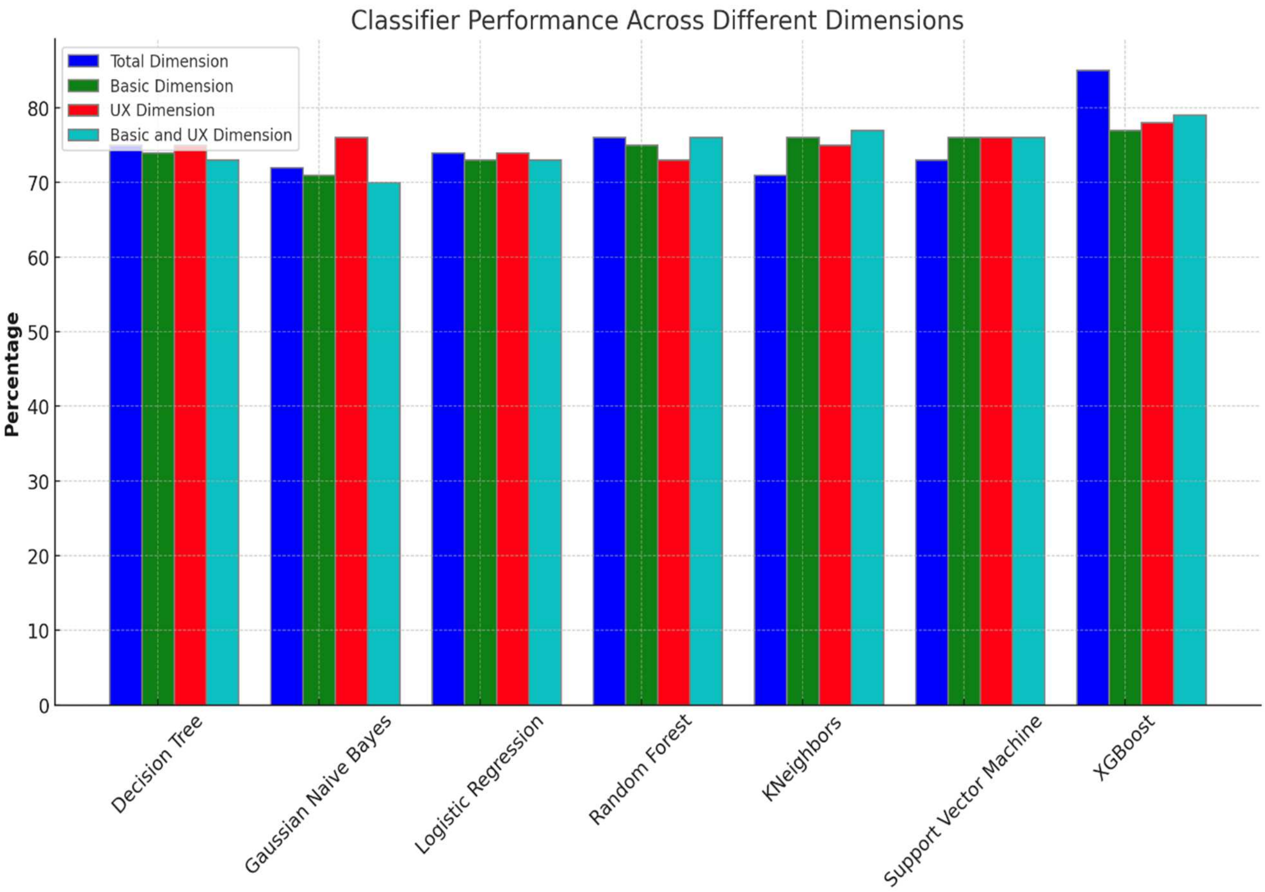

Fig. 1 The performance evaluation of machine learning Metrix

elements of UX, derived from user feedback comments, into our predictive models alongside satisfaction ratings from previous service usage. In the Table 2 and the Figure 1 we illustrate the result for the better understanding.

TABLE II
PERFORMANCE EVOLUTION OF DIFFERENT MACHINE LEARNING ALGORITHM

Word Count (%)	**Classifiers**	**Total Dimension**	**Basic Dimension**	**UX Dimension**	**Basic and UX Dimension**
Top 20%	Decision Tree	75%	74%	75%	73%
	Gaussian Naive Bayes	72%	71%	76%	70%
	Logistic Regression	74%	73%	74%	73%
	Random Forest	76%	75%	73%	76%
	KNeighbors	71%	76%	75%	77%
	Support Vector Machine	73%	76%	76%	76%
	XGBoost	85%	77%	78%	79%

Drawing on a diverse set of classifiers including Decision Tree, Gaussian Naive Bayes, Logistic Regression, Random Forest, KNeighbors, Support Vector Machine, and XGBoost, our analysis yields promising results. Across these classifiers, XGBoost emerges as the top performer with an accuracy of 85% in predicting customers' return visits, followed closely by Random Forest at 76%. Interestingly, the study reveals varying degrees of accuracy among classifiers, with XGBoost showcasing the highest performance and Gaussian Naive Bayes demonstrating a competitive accuracy of 72% we visualize in the Figure 1.

Moreover, our study highlights the importance of considering affective expressions extracted from customer review comments. We find that a higher word count of feedback contributes to increased prediction accuracy, shedding light on the nuanced relationship between sentiment analysis and customer behavior prediction. These findings underscore the significance of understanding customer preferences and sentiments in enhancing service offerings and fostering customer loyalty in the airline industry. While our study provides valuable insights, it also acknowledges the limitations inherent in predictive modeling and calls for further exploration into the complexities of customer decision-making processes.

IV. CONCLUSION

In conclusion, our study represents a significant step forward in understanding the complex dynamics of customer loyalty within the airline industry. By employing a sophisticated machine learning approach coupled with sentiment analysis techniques, we have unraveled invaluable insights into the factors driving customers' intentions to revisit airline services. Through a comprehensive analysis of feedback comments and satisfaction ratings, we have shed light on the nuanced interplay between customer sentiments, satisfaction levels, and the propensity for return visits.

The results of our study are both promising and enlightening. Across a diverse array of machine learning classifiers, XGBoost emerges as the top performer, boasting an impressive accuracy of 85% in predicting customers' return visits. This underscores the efficacy of machine learning algorithms in discerning subtle patterns within customer feedback data, thereby enabling airlines to anticipate and cater to customer needs more effectively.

Furthermore, our study underscores the critical role of affective expressions extracted from customer review comments. We find that a higher word count of feedback correlates positively with increased prediction accuracy, highlighting the importance of delving deeper into the nuances of customer sentiments. By leveraging sentiment analysis techniques, airlines can gain deeper insights into customer preferences and emotions, thereby enhancing service offerings and fostering long-term loyalty.

However, our study is not without its limitations. While machine learning algorithms offer remarkable predictive power, they are inherently limited by the quality and quantity of input data. Moreover, customer behavior is influenced by a myriad of external factors beyond the scope of our analysis, including economic conditions, competitor actions, and global events. As such, our study serves as a starting point for further exploration into the complexities of customer decision-making processes within the airline industry.

Moving forward, there are several avenues for future research. Firstly, incorporating additional data sources such as social media interactions and demographic information could enrich our analysis and provide deeper insights into customer behavior. Secondly, longitudinal studies tracking customer behavior over extended periods could offer a more comprehensive understanding of the factors influencing repeat patronage. Finally, exploring the potential of advanced machine learning techniques such as deep learning could unlock new avenues for predictive modeling and analysis.

In conclusion, our study offers valuable insights into the intricate interplay between customer sentiments, satisfaction levels, and return visit behavior within the airline industry. By leveraging machine learning and sentiment analysis techniques, airlines can make informed decisions and strategic interventions that resonate with their customers, thereby fostering long-term loyalty and sustainable growth in an increasingly competitive marketplace.

REFERENCES

[1] Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The role of emotions in marketing. Journal of the Academy of Marketing Science, 27(2), 184-206.

[2] Reichheld, F. F., & Sasser, W. E. (1990). Zero defections: Quality comes to services. Harvard Business Review, 68(5), 105-111.

[3] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

[4] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).

[5] Cheng, S., Lam, T., & Hsu, C. H. (2006). Negative word-of-mouth communication intention: An application of the theory of planned behavior. Journal of Hospitality & Tourism Research, 30(1), 95-116.

[6] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[7] Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). Survey Methodology (2nd ed.). Wiley.

[8] Han, H., Back, K. J., & Barrett, B. (2009). Influencing factors on restaurant customers' revisit intention: The roles of emotions and switching barriers. International Journal of Hospitality Management, 28(4), 563-572.

[9] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning (2nd ed.). Springer.

[10] Jang, S., & Yi, S. (2017). A study of the influence of customer experience on revisit intention in the restaurant industry: The mediating role of emotions. Journal of Hospitality and Tourism Management, 30, 88-99.

[11] Kim, J. H., & Lee, H. Y. (2011). Customer satisfaction using low-cost carriers. Tourism Management, 32(2), 235-243.

[12] Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence (Vol. 2, pp. 1137-1143).

[13] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.

[14] Matzat, U., Snijders, C., & Reips, U. D. (2014). A survey method to assess the quality of academic research on the internet: Taking stock and looking ahead. Journal of the Association for Information Science and Technology, 65(7), 1400-1414.

[15] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.

[16] Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology, 29(1), 24-54.

[17] Chanpariyavatevong, K., Wipulanusat, W., Champahom, T., Jomnonkwao, S., Chonsalasin, D., & Ratanavaraha, V. (2021). Predicting airline customer loyalty by integrating structural equation modeling and Bayesian networks. Sustainability, 13(13), 7046.

[18] Zins, A. H. (2001). Relative attitudes and commitment in customer loyalty models: Some experiences in the commercial airline industry. International Journal of Service Industry Management, 12(3), 269-294.

[19] Singh, B. (2021). Predicting airline passengers' loyalty using artificial neural network theory. Journal of air transport management, 94, 102080.

[20] Hayadi, B. H., Kim, J. M., Hulliyah, K., & Sukmana, H. T. (2021). Predicting Airline Passenger Satisfaction with Classification Algorithms. International Journal of Informatics and Information Systems, 4(1), 82-94.

[21] Rahim, A. G. (2017). Customer satisfaction and loyalty: A study of interrelationships and effects in the Nigerian domestic airline industry.

[22] Khudhair, H. Y., Jusoh, A., Mardani, A., Nor, K. M., & Streimikiene, D. (2019). Review of scoping studies on service quality, customer satisfaction and customer loyalty in the airline industry. Contemporary Economics, 375-386.

[23] Salah, M., & Abou-Shouk, M. A. (2019). The effect of customer relationship management practices on airline customer loyalty. Journal of Tourism, Heritage & Services Marketing (JTHSM), 5(2), 11-19.

[24] Wong, J. Y., & Chung, P. H. (2008). Retaining passenger loyalty through data mining: a case study of Taiwanese airlines. Transportation Journal, 47(1), 17-29.

[25] Chen, L. S. L., Wang, M. C. H., Cheng, J. M. S., & Kuntjara, H. (2008). Consumer involvement and brand loyalty in services: evidence from the commercial airline industry in Taiwan. International Journal of Services and Standards, 4(4), 437-452.

[26] Khan, U., & Khan, N. (2014). Customer satisfaction in airline industry. International proceedings of economics development and research, 76, 63.

[27] Jain, P. K., Patel, A., Kumari, S., & Pamula, R. (2022). Predicting airline customers' recommendations using qualitative and quantitative contents of online reviews. Multimedia Tools and Applications, 81(5), 6979-6994.

[28] Passikoff, R. (2006). Predicting market success: New ways to measure customer loyalty and engage consumers with your brand. John Wiley & Sons.

[29] Teichert, T., Shehu, E., & von Wartburg, I. (2008). Customer segmentation revisited: The case of the airline industry. Transportation Research Part A: Policy and Practice, 42(1), 227-242.