

Analysis of the Impact of Tourists' Experiences, Review Ratings, and Budget Levels on Travel Recommendation Scores

Yi Wu

School of Economics and Management, Yanbian University, Yanji, China
wuyi01122006@qq.com

Abstract. As travel platforms and agencies increasingly rely on user data for service recommendations, tourist ratings, review sentiments, and budget characteristics have become key indicators for assessing travel experience quality and recommendation effectiveness. Data-driven recommendation systems, by comprehensively analyzing user ratings, sentiment trends, and budget levels, significantly enhance the predictive power of recommendation scores and the precision of tiered recommendations, helping businesses achieve higher conversion rates and user satisfaction. Based on real cross-sectional data, this study conducts a comprehensive analysis using OLS and logistic regression, addressing the gap in the predictive power of the "experience-emotion-budget" framework for recommendation scores and entry into high-recommendation-score groups, while providing actionable management insights and theoretical contributions and providing quantitative evidence for understanding the formation mechanism of tourism recommendation scores. This study finds that subjective experience (scores and emotional responses) and financial capacity (budget level) significantly influence recommendation effectiveness, with their combined predictor strength outperforming that of any single variable.

Keywords: User Ratings, Review Sentiment, Budget Level, Travel Recommendation Score

1. Introduction

As travel platforms and agencies increasingly rely on user data for service recommendations, tourist ratings, review sentiments, and spending patterns have become crucial indicators for assessing travel experience quality and recommendation effectiveness. While prior research has highlighted the importance of data-driven recommendations in tourism services, further analysis is required to identify specific factors significantly influencing travelers' recommendation preferences using concrete user data. This study employs three analytical approaches based on the tourism_dataset.xlsx dataset: First, linear regression analysis using Ordinary Least Squares (OLS) examines the impact of variables such as user ratings, review sentiment, and budget levels on the continuous recommendation score (Recommendation_Score). Second, a logistic regression model was constructed with $\text{Recommendation_Score} \geq 70$ defined as the high-recommendation group to predict participants' inclusion in this category. Finally, 5-fold cross-validation was conducted to

evaluate the model's stability across different data splits. The core explanatory variables were User_Rating, Review_Sentiment, and Budget, while control variables included Travel_Group_Size, Avg_Spending_per_day, Duration, Season, and Location_Type.

The study focuses on analyzing the impact of user ratings, review sentiment, and budget levels on travel recommendation scores.

This paper primarily investigates whether tourist experience evaluations, review sentiments, and budget levels significantly influence tourism recommendation scores, and whether these variables can effectively predict whether tourists fall into the high-recommendation-score group. Based on the data fields and research objectives, the following hypotheses are proposed:

H1: The higher the user rating, the higher the recommendation score.

H2: The more positive the comment sentiment, the higher the recommendation score.

H3: The higher the budget level, the higher the recommended score.

H4: User ratings, review sentiment, and budget levels significantly increase the likelihood of tourists being assigned to the high-recommendation score group.

2. Literature review

Most existing studies treat tourists' subjective experiences, review emotions, and consumption levels in isolation. They look at how each factor sways travel choices or satisfaction, yet rarely pull these threads together to see how they jointly determine recommendation scores.

Sai Liang sees online star ratings as a straightforward numerical gauge of overall service experience and subjective satisfaction. He points out that higher ratings track closely with positive emotional pleasure, meaning better perceived experiences translate into higher scores, and that user ratings and review emotions significantly enhance travel recommendation scores. Liang emphasized that subjective experience is the key determinant of recommendation effectiveness [1].

Turning to review text, Ming Zhang argued that positive emotions reflect approval of accommodation, scenery, and services, while negative emotions indicate dissatisfaction; emotional scores are quantifiable textual indicators independent of star ratings, with more positive reviews correlating with higher recommendation scores, particularly positive emotions being the most influential factor [2]. Xue Li concluded that control variables such as season and location type do not affect ratings' dominant influence on recommendation scores, while Li affirmed the robustness of core explanatory variables [3].

On the spending side, Wei Lu sorts tourists into budget tiers based on disposable travel expenditures—low, medium, or high. He notes that price level colors evaluation criteria: high-budget travelers let minor flaws slide, whereas low-budget travelers watch every penny and punish identical experiences with lower scores. Higher budget tiers correlate with markedly increased recommendation scores and likelihood rates. Lu emphasizes that consumption tier serves as a critical predictive variable [4].

Alyoubi offers a cautionary note. Online star ratings, in his view, are simplified numerical outputs that reflect overall tourist satisfaction, yet single ratings fail to distinguish whether dissatisfaction stems from service quality, scenic appeal, or pricing costs. Their study demonstrates that user ratings only capture comprehensive perceptions but cannot capture specific negative aspects related to budget constraints; they must be combined with multi-dimensional sentiment analysis for accurate recommendation prediction [5]. Shafaei et al. further propose that textual sentiment analysis provides granular supplementary indicators to user ratings: positive sentiment indicates exceeding expectations, negative sentiment suggests falling short of expectations, and neutral sentiment

reflects average experiences; emotional polarity independently influences recommendation behavior beyond rating metrics [6].

With these insights in mind, we turn to the `tourism_dataset.xlsx`. Our aim is to see how user ratings, sentiment, and budget tiers jointly shape recommendation scores. We start with OLS linear regression to estimate the effect of user ratings, review sentiment, and budget level on the continuous `Recommendation_Score`. We then split the sample at 70, defining scores at or above this threshold as the high-recommendation group, and use logistic regression to estimate the probability of landing in that group. Finally, we check whether our results hold up under 50% stratified cross-validation. By grounding the analysis in real tourist data, we hope to show exactly how subjective experiences, review sentiment, and spending levels feed into recommendation preferences, offering concrete evidence on how tourism recommendation scores take shape.

3. Research process

3.1. Data and variables

Our raw data come from the `tourism_dataset.xlsx` file, covering 2,300 tourist records and ten variables in total. These include trip profile information—season, destination type, and duration—alongside experiential indicators such as user ratings, review sentiment, budget tier, daily spending, and the number of travelers.

Table 1 summarizes how each variable is operationalized. `Recommendation_Score` is treated as the dependent variable, capturing tourists' overall willingness to recommend a travel experience. The main predictors are `User_Rating` (the star score given by the tourist), `Review_Sentiment` (negative, neutral, or positive), and `Budget` (low, medium, or high). We also account for several potential confounders: travel group size, average daily expenditure, trip duration, season, and location type.

Algorithm 1: Data loading and variable construction

Input: `tourism_dataset.xlsx`

Output: Cleaned dataset D; `Recommendation_Score`; `High_Recommendation`

1. Read `tourism_dataset.xlsx` into dataset D.

2. Check basic data quality:

number of observations, variables, missing values, duplicated records.

3. Define variables:

`Y_continuous` = `Recommendation_Score`

`X_core` = {`User_Rating`, `Review_Sentiment`, `Budget`}

`X_control` = {`Travel_Group_Size`, `Avg_Spending_per_day`, `Duration`,
`Season`, `Location_Type`}

4. Construct binary outcome for logistic regression:

If `Recommendation_Score` $\geq$ 70:

`High_Recommendation` = 1

Else:

`High_Recommendation` = 0

5. Exclude `User_ID` from all regression models.

...

Return D.

Table 1. Variable definitions table

variable name	variable role	illustration
Recommendation_Score	dependent variable	Measures the overall evaluation of tourists regarding their travel experience or recommended outcomes.
User_Rating	Core Explanatory Variable	Reflects the ratings given by tourists for their travel experience.
Review_Sentiment	Core Explanatory Variable	Classified as negative, neutral, or positive reviews
Budget	Core Explanatory Variable	Categorized into low budget, medium budget, and high budget
Travel_Group_Size	controlled variable	Represents the number of accompanying tourists
Avg_Spending_per_day	controlled variable	Represents the average daily spending level of tourists.
Duration	controlled variable	Classified into short-term, medium-term, and long-term travel
Season / Location_Type	controlled variable	Control for differences in travel contexts and tourist destinations

Table 2. Descriptive statistics table

variable	sample capacity	mean value	standard deviation	least value	median	crest value
User_Rating	2300	3.034	1.157	1.000	3.100	5.000
Avg_Spending_per_day	2300	276.178	129.232	50.110	277.865	499.950
Recommendation_Score	2300	65.785	23.596	0.000	66.000	100.000

Table 2 summarizes the basic statistics. The mean user rating is 3.034, so tourists in this sample typically gave middle-of-the-road scores. Recommendation scores averaged 65.785, with a median of 66.000—both figures point to a reasonably high willingness to recommend. That said, the standard deviation of 23.596 shows considerable spread; not every tourist felt the same way. Average daily spending came in at 276.178, which confirms that the sample covers a broad spectrum of budgets.

3.2. Descriptive analysis

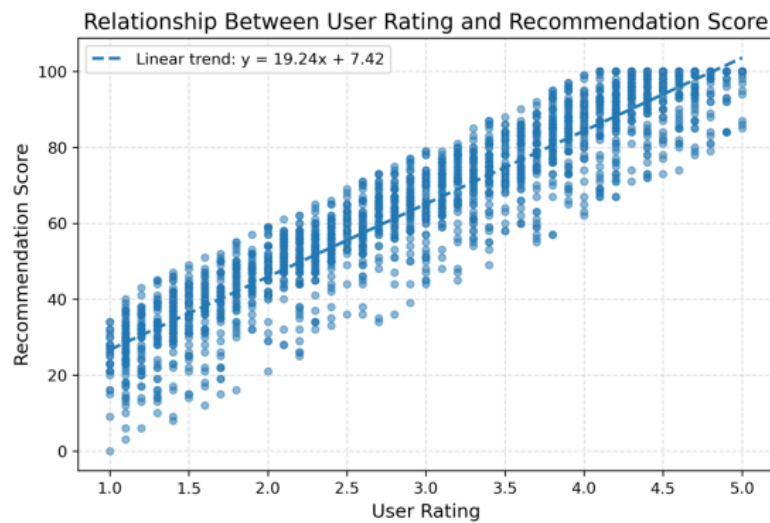


Figure 1. Relationship between user ratings and recommendation scores

Figure 1 shows a clear positive correlation between user ratings and recommendation scores. The linear trend line in the scatter plot is as follows: Recommended Score = $19.24 \times \text{User Rating} + 7.42$

At first glance, tourists who give higher ratings also tend to receive higher recommendation scores. Low-rated trips cluster toward the bottom of the scale with wider dispersion, while high-rated ones sit more tightly in the upper range. This pattern lends initial credence to H1: user ratings do appear to matter for how likely a tourist is to recommend an experience. Whether this link holds once we account for confounders such as budget, sentiment, and trip characteristics is something we test next through OLS regression.

3.3. Regression results

We use OLS regression to model how user ratings, review sentiment, and budget tier shape continuous recommendation scores. Recommendation_Score is the outcome variable, while travel group size, daily spending, trip duration, season, and destination type are included as controls. The estimates are reported in Table 3.

Algorithm 2: OLS regression analysis

Input: Cleaned dataset D

Output: coefficients, t values, p values, R-squared, adjusted R-squared

1. Set dependent variable:

$Y = \text{Recommendation_Score}$.

2. Set predictors:

$X = \{\text{User_Rating}, \text{Review_Sentiment}, \text{Budget}, \text{control variables}\}$.

3. Convert categorical variables into dummy variables:

Review_Sentiment: reference = Negative

Budget: reference = Low

Duration: reference = Short

Season: reference = Spring

Location_Type: reference = Beach

4. Estimate the OLS model:

Recommendation_Score = $\beta_0 + \beta_1 \text{User_Rating}$
+ $\beta_2 \text{Review_Sentiment}$
+ $\beta_3 \text{Budget}$
+ $\beta_4 \text{Control variables} + \epsilon$.
5. Extract model results:
coefficients, t values, p values, R^2 and adjusted R^2 .
6. If $p < 0.05$, treat the variable as statistically significant.
...
Return OLS regression results.

Table 3. Core results of OLS linear regression

variable	coefficient	T value	P value	conclusion
User Rating	19.4332	341.5677	0.0000	Significantly positive
Comment sentiment: Neutral, relatively negative	14.5820	65.9647	0.0000	Significantly positive
Comment sentiment: Positive, relatively negative	19.2345	96.1262	0.0000	Significantly positive
Budget Level: Medium, relatively low budget	4.8966	30.5981	0.0000	Significantly positive
Budget Level: High, relatively low budget	9.2731	56.8151	0.0000	Significantly positive
Number of travelers	-0.0275	-0.5947	0.5521	quiet
Daily average consumption	-0.0001	-0.2107	0.8331	quiet

The OLS model tracks the data closely. With an R^2 of 0.9824 and an adjusted R^2 only a hair lower at 0.9823, it accounts for virtually all of the variation in recommendation scores.

Ratings are the single biggest driver. Raising the user rating by one point pushes the recommendation score up by about 19.43 points ($\beta = 19.4332$, $t = 341.57$, $p < 0.001$), and that relationship stays intact once we control for the rest of the variables.

Review sentiment adds its own punch. Against negative reviews as the baseline, neutral ones lift scores by roughly 14.58 points ($t = 65.96$, $p < 0.001$), while positive ones add about 19.23 points ($t = 96.13$, $p < 0.001$). The pattern is straightforward: as the tone of the review brightens, the recommendation score climbs right along with it.

Budget tier follows a similar graded pattern. Medium-budget travelers score around 4.90 points higher than the low-budget group, and high-budget travelers sit about 9.27 points above them. Both gaps hold up statistically.

Group size and daily spending, by contrast, never reach significance. That suggests a tourist's decision to recommend hinges less on how many people came along or what they shelled out each day, and more on the rating they assigned, the sentiment they expressed, and the budget category they fell into. Put together, these results give clear support to H1, H2, and H3.

3.4. Logistic regression results

After exploring the linear patterns, we switch to a binary setup to identify what pushes tourists into the high-recommendation bracket. We use 70 as the cut-off: scores at or above this threshold are coded 1, and anything below is coded 0. Logistic regression then estimates how each variable affects the probability of landing in that high-score group (see Table 4).

Algorithm 3: Logistic regression and ROC analysis

Input: Dataset D with High_Recommendation

Output: classification indicators, ROC curve and AUC

1. Split D into training set and test set:

training set = 70%, test set = 30%, stratified by High_Recommendation.

2. Define three logistic regression models:

Model 1: High_Recommendation ~ User_Rating

Model 2: High_Recommendation ~ User_Rating + Review_Sentiment + Budget

Model 3: High_Recommendation ~ User_Rating + Review_Sentiment + Budget

+ Travel_Group_Size + Avg_Spending_per_day

+ Duration + Season + Location_Type

3. Fit each model on the training set.

4. Predict $P(\text{High_Recommendation} = 1)$ on the test set.

5. Convert probability into class label:

If predicted probability ≥ 0.5 , predicted class = 1;

otherwise, predicted class = 0.

6. Calculate Accuracy, Precision, Recall, F1-score and AUC.

7. Draw ROC curves for model comparison.

...

Return logistic regression results.

Table 4. Evaluation metrics for the logistic regression model

model	Accuracy	Precision	Recall	F1	AUC
Model 1: Only User Ratings	0.9029	0.8899	0.9037	0.8968	0.9689
Model 2: Add comment sentiment and budget	0.9551	0.9709	0.9317	0.9509	0.9958
Model 3: Complete Model	0.9464	0.9582	0.9255	0.9415	0.9950

All three logistic models classify the test data well. Model 1, built only on user ratings, reaches 0.9029 accuracy and 0.9689 AUC—already enough to tell high- from low-recommendation tourists apart. Once review sentiment and budget enter the equation in Model 2, accuracy climbs to 0.9551 and AUC to 0.9958, a jump that confirms these variables add genuine predictive value. The full model records 0.9464 accuracy and 0.9950 AUC, essentially matching Model 2 and suggesting that the core set of rating, sentiment, and budget captures most of what drives the classification.

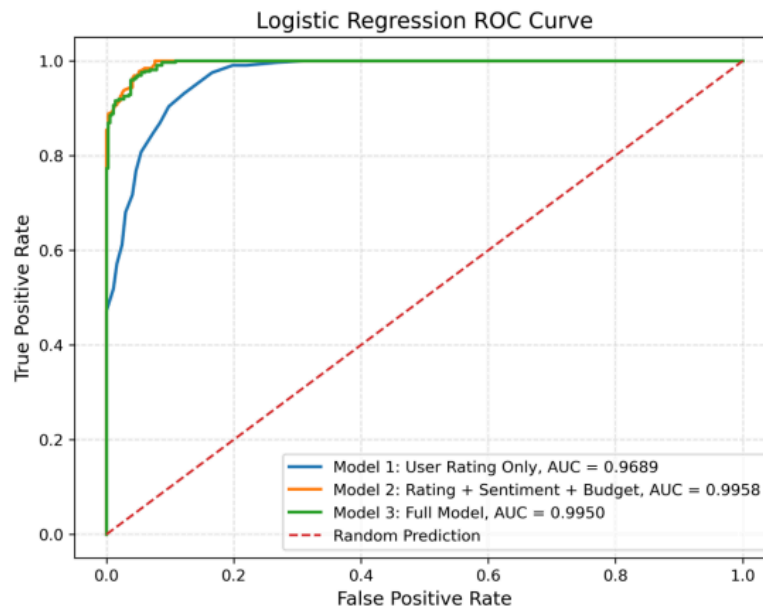


Figure 2. ROC curve of the logistic regression model

Figure 2 traces the ROC curves for all three specifications. Each curve lies far above the random-guess diagonal, confirming that the logistic models discriminate well between high- and low-recommendation tourists. Model 2 and the full model both push the AUC almost to 1, suggesting that adding sentiment and budget sharply improves our ability to spot travelers who will land in the high-recommendation group. These findings line up with H4: ratings, sentiment, and budget all raise the odds of receiving a high recommendation score.

3.5. K-fold cross validation

To check whether the logistic model holds up, we used 5-fold stratified cross-validation. We carved the data into five roughly equal pieces, trained on four, and tested on the leftover one, rotating through all five so each chunk served as the test set once. This smooths out the chance variation that comes from relying on a single split (see Table 5).

Algorithm 4: Five-fold cross validation and robustness checks

Input: Cleaned dataset D

Output: cross-validation metrics and robustness results

Part A: Five-fold cross validation

1. Use High_Recommendation as the binary outcome.
2. Standardize numerical variables and convert categorical variables into dummy variables.
3. Split D into 5 stratified folds.
4. For each fold $i = 1, \dots, 5$:
Use fold i as validation set.
Use the remaining folds as training set.
Fit the logistic regression model.
Predict validation-set probabilities.
Calculate Accuracy, F1-score and AUC.
5. Average performance indicators across the 5 folds.
6. Combine out-of-fold predictions and draw OOF ROC curve.

...

Part B: Robustness checks

7. Re-estimate the OLS model using HC3 robust standard errors.

8. Compare ordinary OLS and HC3 results:

Check whether core variables keep the same direction and significance.

9. Repeat logistic regression under alternative thresholds:

Thresholds = {65, 70, 75}.

10. If model performance remains high and core variables remain significant, the results are considered robust.

...

Return validation and robustness results.

Table 5. Cross-validation results for 50% fold cross-validation

model	Accuracy mean	F1 mean	AUC mean
Model 1: Only User Ratings	0.9157	0.9100	0.9767
Model 2: Add comment sentiment and budget	0.9643	0.9617	0.9969
Model 3: Complete Model	0.9630	0.9600	0.9967

At the 50% confidence level, Model 1 posts a mean AUC of 0.9767, confirming that user ratings alone carry strong predictive power. Model 2, with sentiment and budget added, lifts mean accuracy to 0.9643 and mean AUC to 0.9969—the best result of the three. The full model registers 0.9630 accuracy and 0.9967 AUC, virtually neck-and-neck with Model 2, so the extra control variables do not add much beyond what the core predictors already capture.

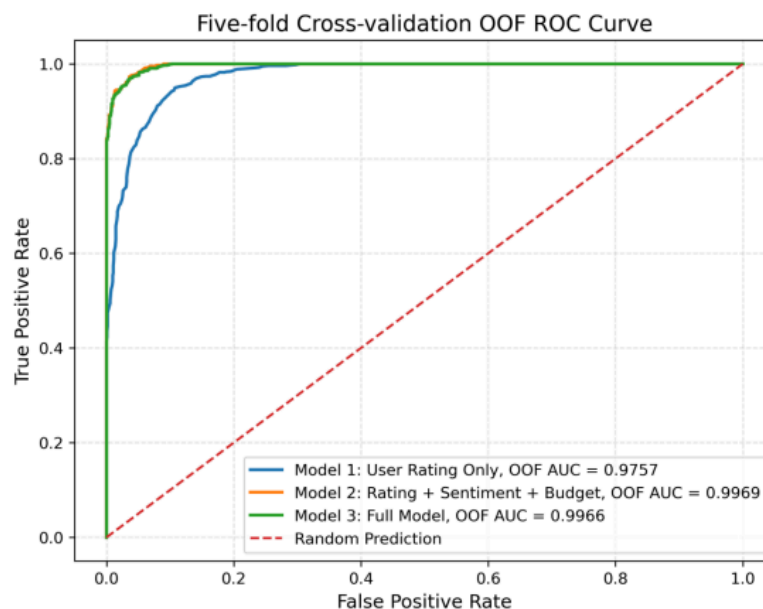


Figure 3. Cross-validation OOF ROC curve with a 50% fold ratio

The out-of-fold ROC curves in Figure 3 tell the same story: Model 2 and the full model both keep AUC values near 1. This means the strong results are not an artifact of one lucky train-test split; they

replicate cleanly across different folds. The k-fold check therefore reinforces the logistic regression findings we presented above.

3.6. OLS robust standard error test

Table 6. Results of the robust standard error test for OLS

variable	OLS coefficient	OLS p value	HC3 Stable Standard Error	HC3 p value	Robustness Conclusion
User Rating	19.4332	0.0000	0.0631	0.0000	Robustly significant
Comment sentiment: Neutral, relatively negative	14.5820	0.0000	0.2146	0.0000	Robustly significant
Comment sentiment: Positive, relatively negative	19.2345	0.0000	0.1957	0.0000	Robustly significant
Budget Level: Medium, relatively low budget	4.8966	0.0000	0.1505	0.0000	Robustly significant
Budget Level: High, relatively low budget	9.2731	0.0000	0.1694	0.0000	Robustly significant
Number of travelers	-0.0275	0.5521	0.0466	0.5560	Still not significant
Daily average consumption	-0.0001	0.8331	0.0005	0.8365	Still not significant

Table 6 reruns the OLS estimates with HC3 robust standard errors. The main variables—user ratings, neutral and positive review sentiment, medium and high budget tiers—all keep their positive signs and remain strongly significant ($p < 0.001$). Switching to robust standard errors does not change the inference for any of these predictors. Meanwhile, the number of travelers and daily average spending stay insignificant under both conventional and robust estimation. These checks suggest that heteroskedasticity is not distorting our results; the core findings hold up.

4. Conclusion

This paper primarily investigates whether user ratings, review sentiment, and budget levels are key factors influencing travel recommendation scores. Among these, user rating is the most critical variable: higher ratings correspond to higher recommendation scores; more positive review sentiments correlate with stronger tourist recommendation tendencies; and higher budget levels increase both recommendation scores and the probability of falling into the high-recommendation score category.

Theoretically, this study uses real tourism data to demonstrate how tourists' subjective experiences, review sentiments, and spending levels influence recommendation preferences, providing quantitative evidence for understanding the formation mechanism of travel recommendation scores. From a management perspective, tourism platforms and travel agencies should prioritize monitoring tourist ratings and review sentiments, promptly identifying negative feedback, and implementing service improvements. Additionally, differentiated recommendation

strategies can be developed based on budget tiers: emphasizing service quality and personalized experiences for high-budget travelers, while highlighting cost-effectiveness and stable basic services for low-budget travelers.

This study also has certain limitations. First, the data are cross-sectional and can only illustrate statistical relationships between variables, but cannot rigorously establish causal links. Second, the dataset lacks more detailed user profile variables such as age, income, travel purpose, and repurchase behavior. Future research could integrate real transaction data, long-term tracking information, and richer user characteristics to further analyze the relationship between travel recommendation scores and actual purchasing behavior.

References

- [1] Liang, S. (2023). Research on online tourist rating sentiment heterogeneity and influencing factors based on three-dimensional theory. *Nankai Business Review*, 26(4), 223–233.
- [2] Zhang, M. (2025). Analysis of tourist experience and emotional tendency of rural homestays: A text mining study based on Ctrip reviews. *Tourism Development Frontier*, 3, 45–58.
- [3] Li, X. (2026). Research on tourist satisfaction of Wuzhen scenic spot based on online reviews. *Tourism Forum*, 1, 78–89.
- [4] Lu, W. (2025). Why do tourists give more generous ratings online? Tourist bias in online reviews. *Information Systems Research*, 36(2), 512–530.
- [5] Alyoubi, A. (2025). Multi-aspect sentiment classification of Arabic tourism reviews using BERT and classical machine learning. *Journal of Sensor and Actuator Networks*, 14(3), 42.
- [6] Shafaei, A., Li, M., & Wang, H. (2025). Towards a better understanding of tourist satisfaction in ethnic-minority villages: A sentiment analysis approach. *Journal of Vacation Marketing*, 31(2), 189–203.