

Tourist Satisfaction Analysis of East Aceh Coastal Ecotourism Using Stacking Machine Learning and GUI Integration

Imam Gunanjar¹, Ryanda Saputra², Shary Armonitha Lusiana³, Martahadi Mardhani⁴,
M. Khairul Anam⁵

¹Department of Management, Universitas Samudra, Indonesia

²Department of Accounting, Universitas Samudra, Indonesia

³Department of Informatics Engineering, Universitas Putra Indonesia YPTK Padang, Indonesia

⁴Department of Developmet Economics, Universitas Samudra, Indonesia

⁵Department of Informatics, Universitas Samudra, Indonesia

Article Info

Article history:

Received May 09, 2026

Revised June 29, 2026

Accepted July 26, 2026

Keywords:

Tourist Satisfaction

Coastal Ecotourism

Stacking

Machine Learning

GUI

ABSTRACT

Tourist satisfaction is a key indicator for evaluating service quality and improving destination management in coastal ecotourism. East Aceh has several coastal tourism destinations with strong ecotourism potential, but tourist opinions are not always systematically analyzed for service improvement. This study analyzes tourist satisfaction in East Aceh coastal ecotourism using a stacking machine learning approach and integrates the best-performing model into a Graphical User Interface (GUI). The dataset consists of 1,679 tourist opinions collected from several coastal tourism objects in East Aceh and labeled into three classes: Satisfied, Neutral, and Dissatisfied. To prevent data leakage, the dataset was first divided into training and testing sets using a 60:40 split; TF-IDF fitting and SMOTE resampling were then applied only to the training data, while the testing set was kept unchanged. The evaluated base models include Naive Bayes, Random Forest, Support Vector Machine, and XGBoost, with Logistic Regression used as the stacking meta-model. The results show that the stacking model with SMOTE achieved the best performance with 87.05% accuracy, 87% precision, 87% recall, and 87% F1-score, outperforming the individual models in this experimental setting. However, the improvement over stacking without SMOTE was modest and was not statistically tested. The best-performing model was implemented in a GUI-based prototype that demonstrates how tourist opinions can be uploaded and classified into satisfaction categories. The prototype provides a preliminary interface for presenting classification results and has not yet been validated as an operational decision-support system.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Imam Gunanjar,

Department of Management, Universitas Samudra,

Jl. Prof. Dr. Syarief Thayeb, Langsa City, Aceh 24416, Indonesia.

Email: imamgunanjar@unsam.ac.id

1. INTRODUCTION

Ecotourism-based coastal tourism plays an important role in promoting local economic growth while supporting environmental conservation [1]. The coastal areas of Aceh, particularly the eastern region covering East Aceh Regency, Langsa City, and Aceh Tamiang Regency, have various ecotourism potentials, including mangrove forests, beaches, coastal landscapes, and other natural destinations that can attract both domestic and international tourists [2]. These potentials can positively contribute to surrounding communities through increased economic activities, such as transportation services, culinary businesses, local trade, micro-enterprises, and other tourism-

supporting services [3]. However, these ecotourism potentials will not provide optimal benefits if service quality and tourist experience are not managed properly, consistently, and sustainably [4].

Tourist satisfaction is one of the key indicators for evaluating the success of ecotourism destination management [5]. Satisfied tourists are more likely to revisit, provide positive reviews, and recommend the destination to others [6]. Conversely, tourist dissatisfaction may reduce the destination's image and decrease future visitation interest [7]. In the context of coastal ecotourism, tourist satisfaction can be influenced by various aspects, such as environmental cleanliness, accessibility, supporting facilities, service quality, safety, comfort, and the availability of tourism information [8]. Therefore, tourist satisfaction analysis is necessary as a basis for destination managers and local governments to formulate more targeted service improvement strategies [9].

In the digital era, tourists increasingly share their experiences and evaluations through various platforms, including reviews, comments, and survey data. These data are valuable because they represent tourists' perceptions of the services they receive. However, conventional satisfaction analysis often faces limitations, especially when dealing with large, diverse, and complex datasets. Manual analysis also requires considerable time and may lead to subjective interpretation. Therefore, a data-driven approach is needed to process tourist satisfaction information more quickly, accurately, and consistently.

Machine learning has become a widely used approach for analyzing tourist reviews and satisfaction [10]. Through classification techniques, machine learning can help identify tourist satisfaction tendencies into specific categories, such as satisfied and dissatisfied [11]. Several previous studies have demonstrated the effectiveness of machine learning in sentiment and satisfaction analysis in the tourism sector. Study [12], analyzed tourist attraction sentiment in the Special Region of Yogyakarta based on Google Maps reviews using Support Vector Machine with TF-IDF features and achieved an accuracy of 83.8%. Study [13], examined sentiment toward waterfall tourism in Central Lombok using Google Maps reviews, automatic labeling based on IndoBERT, and several SVM variants, where LinearSVC achieved the best performance with 84% accuracy and an F1-score of 86%.

Other studies have also shown that machine learning algorithms can provide strong performance in text-based data analysis. Study [14], applied Naive Bayes to analyze sentiment in Instagram comments from BPS and achieved an accuracy of 94%, although the study still faced challenges related to highly imbalanced class distribution. Study [15], compared Naive Bayes and SVM on tourist reviews of Alun-Alun Kota Batam and found that SVM outperformed Naive Bayes, achieving 94% accuracy compared to 83%. Meanwhile, [16], implemented a deep learning approach using Bi-LSTM and Word2Vec on Bali tourism reviews and obtained an accuracy of 96.86%. These findings indicate that machine learning and deep learning approaches have strong potential for analyzing tourist perceptions. However, most previous studies still focused on using a single algorithm or comparing several models separately.

Although previous studies have reported promising results, the use of a single algorithm is often insufficient to capture complex data patterns. Each algorithm has different characteristics. Support Vector Machine is known for its ability to handle high-dimensional data [17], Naive Bayes is effective for simple and fast text processing [18], Random Forest can capture nonlinear relationships through tree-based ensemble mechanisms [19], and XGBoost offers strong optimization capability for improving classification performance [20]. These different characteristics suggest that combining multiple algorithms may produce a more stable and accurate model than relying on a single classifier.

Based on this background, this study proposes a stacking machine learning approach to analyze tourist satisfaction in Aceh coastal ecotourism. The proposed model combines Support Vector Machine, Naive Bayes, Random Forest, and XGBoost as base learners, with Logistic Regression used as the meta-model. Through the stacking scheme, predictions from multiple base models are combined to generate a more optimal final decision [21]. This approach is expected to classify tourist satisfaction more accurately, particularly in distinguishing between satisfied and dissatisfied categories based on the available data.

In addition to developing a classification model, this study also integrates the analysis results into a Graphical User Interface (GUI). GUI integration is important because machine

learning results should not only demonstrate good performance but also be practically usable by destination managers and stakeholders. Through the GUI, users can input data, obtain tourist satisfaction classification results, and use the information as a basis for determining service improvement priorities. Thus, this study does not only focus on technical modeling aspects but also on the development of an accessible analytical system to support data-driven decision-making.

The novelty of this study lies in the application of stacking machine learning for analyzing tourist satisfaction in Aceh coastal ecotourism, covering East Aceh Regency, Langsa City, and Aceh Tamiang Regency. Unlike previous studies that generally used a single algorithm or compared multiple algorithms separately, this study combines several algorithms within an ensemble stacking scheme and integrates the results into a GUI. Through this approach, the study is expected to contribute to the development of a more accurate, practical, and applicable tourist satisfaction analysis system to support service quality improvement in Aceh coastal ecotourism.

Based on the above discussion, this study aims to analyze tourist satisfaction in Aceh coastal ecotourism using a stacking machine learning approach, evaluate model performance based on accuracy, precision, recall, and F1-score, and integrate the model into a GUI as an analytical tool for destination managers. The results of this study are expected to help managers and stakeholders understand tourist satisfaction patterns, identify service aspects that require improvement, and formulate more targeted service enhancement strategies based on data analysis.

2. RESEARCH METHOD

This study was conducted through several main stages, including problem identification, literature review, tourist opinion data collection, data labeling, class imbalance handling using SMOTE, preprocessing, feature extraction using TF-IDF, stacking machine learning model development, model evaluation, and system testing through a Graphical User Interface (GUI). These stages were designed to produce a tourist satisfaction classification model that is not only computationally accurate but also practically usable by managers of coastal ecotourism destinations in East Aceh.

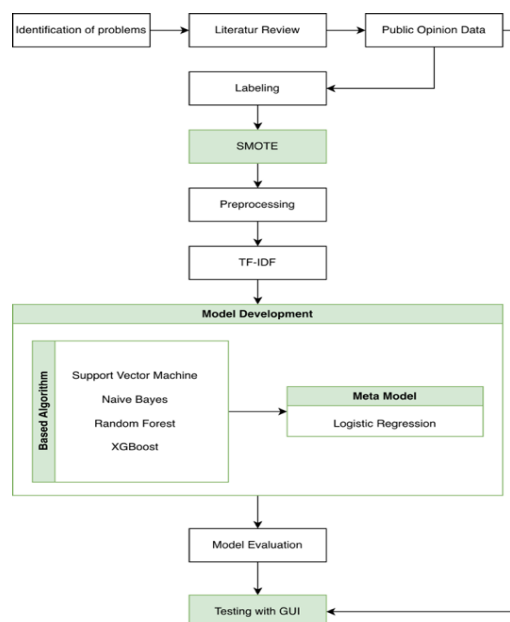


Figure 1. Methodology Flow

2.1. Tourist Opinion Data

The dataset used in this study consisted of 1,679 tourist reviews collected from Google Reviews. The reviews were obtained from nine coastal tourism destinations located in the eastern coastal region of Aceh. These destinations were distributed across three administrative areas, namely Langsa City, Aceh Tamiang Regency, and East Aceh Regency, with three tourism destinations selected from each area. The dataset contained tourists' opinions, comments, and

reviews that reflected their experiences after visiting the destinations, including both positive and negative experiences. These reviews were used as the primary data source because they represent real user-generated feedback from visitors.

2.2. Data Labeling

After the tourist review data were collected, the next stage was data labeling. The initial labeling process was conducted manually by one researcher to assign each review into one of three satisfaction categories: Satisfied, Neutral, and Dissatisfied. Each label was determined based on the meaning, sentiment, and contextual interpretation expressed in the review text. Because the initial labeling was performed by a single researcher, inter-annotator agreement, such as Cohen's kappa, was not calculated in this study.

After the initial labeling process was completed, the labeling results were consulted with relevant experts to improve the appropriateness of the assigned categories. This consultation involved economic experts and representatives of the East Aceh Tourism Office, who provided feedback based on the practical context of tourism services and visitor satisfaction. If a label was considered less appropriate, it was reviewed and revised based on expert feedback. This process improved contextual validity, although it does not replace formal reliability testing using multiple independent annotators.

2.3. Class Imbalance Handling Using SMOTE

After the labeling process, the dataset was examined to determine whether there was an imbalance in the number of opinions in each class. The original distribution consisted of 748 Satisfied opinions, 497 Neutral opinions, and 434 Dissatisfied opinions. The dataset was then divided into training and testing sets using a 60:40 split before any resampling process was applied. This procedure was used to ensure that the testing data remained original and unseen during model training.

The training set consisted of 508 Satisfied reviews, 262 Neutral reviews, and 237 Dissatisfied reviews. SMOTE was applied only to the TF-IDF vectors in the training set, not to the full dataset and not to the raw text. Therefore, the resampled training data contained approximately 508 samples per class, following the largest class in the training subset. The value of 748 refers only to the number of Satisfied reviews in the original complete dataset and is not the number of samples produced by SMOTE after the 60:40 split. This clarification is important to avoid misunderstanding and to prevent possible data leakage [22].

The no-resampling baseline, SMOTE configuration, and class-weighting configuration were evaluated using the same unchanged testing set. Class weighting was evaluated for algorithms that support weighted learning, namely Random Forest, Support Vector Machine, XGBoost, and the stacking ensemble. Naïve Bayes was not evaluated using class weighting because the conventional Multinomial Naïve Bayes implementation does not directly support the same `class_weight` mechanism as the other classifiers. Model performance was compared using accuracy, macro-averaged precision, macro-averaged recall, and macro-averaged F1-score [23].

2.4. Data Preprocessing

The preprocessing stage was conducted to clean and prepare text data before feature extraction. Tourist opinion data usually contain irrelevant elements, such as excessive punctuation, numbers, symbols, inconsistent capitalization, informal words, or special characters. If these elements are not removed, they may interfere with the learning process of the model.

Preprocessing included several stages, such as case folding, cleaning, tokenizing, stopwords removal, and stemming. Case folding was used to convert all text into lowercase [24]. Cleaning was performed to remove unnecessary symbols, numbers, and characters [25]. Tokenizing was used to split sentences into individual words [26]. Stopword removal was applied to remove common words that do not have significant meaning in classification [27]. Stemming was used to convert words with affixes into their root forms. The output of this stage was cleaner text data ready for feature extraction [28].

2.5. Feature Extraction Using TF-IDF

After preprocessing, the next step was feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF was used to convert textual data into numerical representations that can be processed by machine learning algorithms [29].

TF-IDF works by calculating the importance of a word in a document based on its frequency in the document and its uniqueness across all documents [30]. Words that frequently appear in one opinion but are not common across the entire dataset receive higher weights. Through this method, important words that help distinguish Satisfied, Neutral, and Dissatisfied opinions can be represented more effectively [31]. The output of the TF-IDF stage was a feature matrix used as input for model development.

2.6. Stacking Machine Learning Model Development

Model development was the main stage of this study. At this stage, the classification model was built using a stacking machine learning approach. Stacking is an ensemble learning method that combines several base models to generate a final prediction through a meta-model [32].

In this study, the base learners consisted of Support Vector Machine, Naïve Bayes, Random Forest, and XGBoost. These algorithms were selected because they have different characteristics. Support Vector Machine is effective for high-dimensional data, Naïve Bayes is simple and efficient for text classification, Random Forest can handle nonlinear patterns, and XGBoost has strong optimization capability for improving classification performance.

The prediction outputs from all base learners were then passed to Logistic Regression as the meta-model. Logistic Regression learned the prediction patterns produced by the base models and generated the final decision, namely whether a tourist opinion belonged to the Satisfied, Neutral, or Dissatisfied category [33]. With the stacking approach, the model was expected to produce more stable and accurate performance compared to using a single algorithm.

2.7. Model Evaluation

After the model was developed, the next stage was model evaluation. Evaluation was conducted to measure the model's ability to classify tourist satisfaction. The evaluation metrics used in this study included accuracy, precision, recall, and F1-score.

Accuracy was used to measure the percentage of correct predictions from all testing data. Precision showed how accurate the model was when predicting a certain class. Recall indicated the model's ability to identify all data that actually belonged to a certain class. Meanwhile, F1-score represented the harmonic mean of precision and recall, making it suitable for evaluating model performance, especially when dealing with class distribution differences.

In addition to these metrics, evaluation was also conducted using a confusion matrix to observe the number of correct and incorrect predictions for each class. The evaluation results were used to determine whether the developed stacking model could provide good performance in classifying tourist satisfaction.

2.8. GUI-Based Prototype testing

The final stage involved functional testing of a GUI-based prototype. The prototype was developed to demonstrate how the trained stacking model can be accessed through a simpler interface without requiring users to interact directly with the machine learning code.

Through the prototype, users can upload tourist opinion data in .xlsx format. The uploaded opinions are then processed by the trained model, and the predicted satisfaction categories are displayed as Satisfied, Neutral, or Dissatisfied. The prototype also presents a visual summary of the classification results to facilitate an initial interpretation of the predicted satisfaction distribution.

The testing conducted in this study was limited to verifying the main functions of the prototype, including file upload, data processing, prediction generation, and result visualization. The prototype has not yet undergone usability testing, user acceptance testing, or operational validation involving destination managers. Therefore, it should be interpreted as a preliminary implementation of the classification model rather than as a validated decision-support system.

3. RESULTS AND DISCUSSION

3.1. Evaluation Result of the Stacking Model

The performance of the proposed stacking model was evaluated using a testing dataset consisting of 672 tourist review data. The dataset used in this study originally contained 1,679 reviews, which were divided into three satisfaction categories: Satisfied, Neutral, and Dissatisfied. In the model implementation, these categories were encoded as follows: Class 0 represents Satisfied, Class 1 represents Neutral, and Class 2 represents Dissatisfied.

The dataset was divided into training and testing sets using a 60:40 train-test split. As a result, 1,007 data were used for training and 672 data were used for testing. The testing set consisted of 240 Satisfied reviews, 235 Neutral reviews, and 197 Dissatisfied reviews. The original class distribution consisted of 748 Satisfied reviews, 497 Neutral reviews, and 434 Dissatisfied reviews. Therefore, the remaining data were used as the training set, consisting of 508 Satisfied reviews, 262 Neutral reviews, and 237 Dissatisfied reviews. This explanation is provided to clarify the difference between the original class distribution and the number of samples reported in the testing set.

To build the stacking model, several base learners were first trained on the training data. Out-of-fold predictions were generated using k-fold cross-validation on the training set. In each fold, the base models were trained using the training portion of the fold and then used to predict the validation portion. The predictions from all validation folds were then combined to form a new feature set for training the meta-model. In this study, Logistic Regression was used as the meta-model. After the meta-model was trained using the out-of-fold predictions, the base models were retrained using the complete training set and then used to generate predictions on the testing set. These testing predictions were used as input features for the meta-model to produce the final classification results.

The evaluation results are presented using a confusion matrix and a classification report. The confusion matrix was used to observe the number of correct and incorrect predictions for each satisfaction category, while the classification report was used to analyze the model performance based on precision, recall, F1-score, and support.

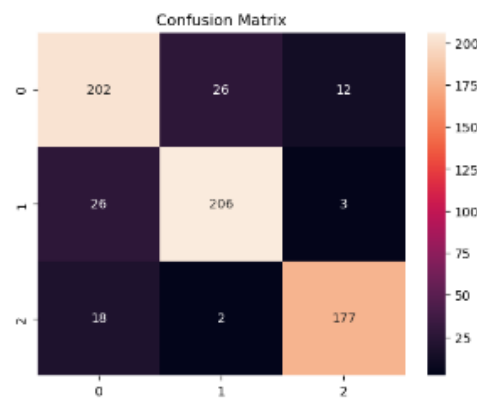


Figure 2. Confusion Matrix of the Stacking Model

Figure 2 shows the confusion matrix of the stacking model. Based on the figure, the stacking model was able to classify most of the testing data correctly. The diagonal values represent the number of correctly predicted data for each satisfaction category. For the Satisfied class, the model correctly classified 202 reviews from a total of 240 actual Satisfied reviews. However, 26 Satisfied reviews were incorrectly classified as Neutral, and 12 reviews were incorrectly classified as Dissatisfied. This indicates that although the model was able to recognize most Satisfied reviews, some reviews still had expressions or word patterns that were similar to the other categories.

For the Neutral class, the model correctly classified 206 reviews from a total of 235 actual Neutral reviews. The remaining errors consisted of 26 Neutral reviews classified as Satisfied and 3 reviews classified as Dissatisfied. This result shows that the model had a strong ability to identify

Neutral reviews, although some overlap still occurred, particularly between Neutral and Satisfied reviews. For the Dissatisfied class, the model produced the best classification result. A total of 177 reviews were correctly classified from 197 actual Dissatisfied reviews. Only 18 Dissatisfied reviews were incorrectly classified as Satisfied, while 2 reviews were classified as Neutral. This result indicates that Dissatisfied reviews had more distinguishable textual patterns compared to the other classes, allowing the stacking model to recognize them more accurately.

Overall, the confusion matrix shows that the model correctly classified 585 out of 672 testing data. This produced an accuracy of approximately 87.05%. The high number of correct predictions on the diagonal indicates that the stacking approach was effective in recognizing the patterns of tourist satisfaction reviews. However, the misclassification between Satisfied and Neutral reviews suggests that some reviews in these two categories may contain similar expressions, making them more difficult for the model to distinguish.

Table 1. Classification Report of the Stacking Model

Class	Precision	Recall	F1-Score	Support
0	0.82	0.84	0.83	240
1	0.88	0.88	0.88	235
2	0.92	0.90	0.91	197
Accuracy			0.87	672
Macro Avg	0.87	0.87	0.87	672
Weighted Avg	0.87	0.87	0.87	672
Overall Accuracy: 0.8705357142857143 or approximately 87.05%.				

Table 1 presents the classification report of the stacking model. The classification report provides a more detailed evaluation based on precision, recall, F1-score, and support. The model achieved an overall accuracy of 0.8705, or approximately 87.05%. This result indicates that the stacking model performed well in classifying tourist reviews into the three satisfaction categories.

For the Satisfied class, the model obtained a precision value of 0.82, recall of 0.84, and F1-score of 0.83 with 240 testing samples. The precision value indicates that 82% of the reviews predicted as Satisfied were correct. Meanwhile, the recall value shows that the model successfully identified 84% of all actual Satisfied reviews. The F1-score of 0.83 indicates that the model had reasonably balanced performance in detecting Satisfied reviews, although some misclassifications still occurred. For the Neutral class, the model achieved a precision of 0.88, recall of 0.88, and F1-score of 0.88 with 235 testing samples. These values show that the model performed consistently in predicting and identifying Neutral reviews. The equal precision and recall values indicate that the model had a balanced ability to avoid false predictions and detect actual Neutral reviews.

For the Dissatisfied class, the model obtained the highest performance compared to the other classes, with a precision of 0.92, recall of 0.90, and F1-score of 0.91 from 197 testing samples. This result shows that the model was highly effective in recognizing Dissatisfied reviews. The precision value of 0.92 indicates that most predictions classified as Dissatisfied were correct, while the recall value of 0.90 shows that the model successfully detected most of the actual Dissatisfied reviews. The macro average values for precision, recall, and F1-score were all 0.87. This indicates that the model achieved balanced performance across all satisfaction categories without being overly biased toward one particular class. The weighted average values were also 0.87, showing that the model maintained stable performance even when considering the different number of testing samples in each class.

Overall, the results from both the confusion matrix and classification report demonstrate that the proposed stacking model was able to classify tourist satisfaction reviews with good performance. The combination of several base learners and Logistic Regression as the meta-model helped produce stable predictions. These results indicate that the stacking approach is suitable for supporting tourist satisfaction analysis in coastal ecotourism, especially as a basis for identifying service quality patterns and supporting data-driven service improvement.

3.2. Comparison of Base Model and Stacking Model

After evaluating the stacking model using the confusion matrix and classification report, the next analysis was conducted by comparing the performance of the proposed stacking model with several individual base models. The evaluated base models consisted of Naïve Bayes, Random Forest, Support Vector Machine, and XGBoost. The comparison was performed using four evaluation metrics, namely accuracy, precision, recall, and F1-score. This comparison aims to determine whether the stacking approach can improve classification performance compared to the use of a single machine learning algorithm.

Table 2. Performance Comparison of Base Models and Stacking Model

Algorithm	Accuracy	Precision	Recall	F1-Score
Naïve Bayes + SMOTE	81.84%	82%	82%	82%
Random Forest + SMOTE	81.69%	82%	82%	82%
Support Vector Machine + SMOTE	85.86%	86%	86%	86%
XGBoost + SMOTE	79.76%	80%	80%	80%
Stacking + SMOTE	87.05%	87%	87%	87%
Naïve Bayes	82.90%	83%	83%	83%
Random Forest	78.80%	79%	79%	79%
Support Vector Machine	80.20%	80%	80%	80%
XGBoost	75.69%	77%	76%	77%
Stacking	85.80%	87%	86%	86%
Random Forest + Class Weighting	80.78%	81%	81%	81%
Support Vector Machine+ Class Weighting	79.71%	80%	80%	80%
XGBoost + Class Weighting	78.32%	78%	78%	78%
Stacking + Class Weighting	83.34%	83%	83%	83%

Table 2 shows the performance comparison of individual machine learning models and the stacking ensemble under three settings: without imbalance handling, with SMOTE, and with class weighting. Overall, the best result was achieved by Stacking + SMOTE, with an accuracy, precision, recall, and F1-score of 87%. This indicates that combining several base learners through stacking and applying SMOTE produced the strongest classification performance in this experiment.

Among the individual models, Support Vector Machine + SMOTE achieved the best performance, with an accuracy of 85.86% and an F1-score of 86%. This shows that SVM benefited the most from SMOTE compared with its original version, which only achieved 80.20% accuracy. Random Forest and XGBoost also improved after SMOTE was applied, increasing from 78.80% to 81.69% and from 75.69% to 79.76%, respectively. However, Naïve Bayes performed slightly better without SMOTE, achieving 82.90%, compared with 81.84% when SMOTE was applied. This suggests that synthetic TF-IDF vectors may not always benefit probabilistic models such as Naïve Bayes.

Class weighting was also tested as an alternative imbalance-handling strategy. The results show that Stacking + Class Weighting achieved 83.34%, while Random Forest, SVM, and XGBoost with class weighting achieved 80.78%, 79.71%, and 78.32%, respectively. These results indicate that class weighting improved some models compared with their original versions, but its performance was generally lower than SMOTE. Therefore, in this experiment, SMOTE was more effective than class weighting for improving classification performance.

It should also be noted that Naïve Bayes was not evaluated using class weighting because the conventional Naïve Bayes implementation does not directly support the same class_weight mechanism as Random Forest, SVM, XGBoost, or stacking-based models. Therefore, Naïve Bayes was only compared in two settings: without SMOTE and with SMOTE.

Overall, the comparison confirms that the stacking ensemble produced the most stable and effective result, especially when combined with SMOTE. Although the improvement from Stacking without SMOTE to Stacking + SMOTE was relatively small, increasing from 85.80% to 87.05%, the experimental procedure was designed to avoid data leakage. The dataset was first divided into training and testing sets using a 60:40 split before any resampling process was applied. TF-IDF fitting and SMOTE oversampling were then performed only on the training set, while the testing set was kept unchanged and was not involved in feature fitting or resampling. Therefore, no information from the testing data or synthetic samples derived from the testing data was used

during model training. However, because no statistical significance test was conducted, the improvement should be interpreted as an empirical gain observed in this experimental setting rather than definitive evidence that SMOTE is statistically superior.

3.3. GUI Implementation for Tourist Satisfaction Prediction

In addition to evaluating the model performance through accuracy, precision, recall, and F1-score, this study also implemented the proposed model into a Graphical User Interface (GUI). The GUI was developed to make the tourist satisfaction classification model easier to use by destination managers or related stakeholders. Through this interface, users can upload tourist opinion datasets in .xlsx format for each tourism destination, process the uploaded data, and obtain prediction results automatically.

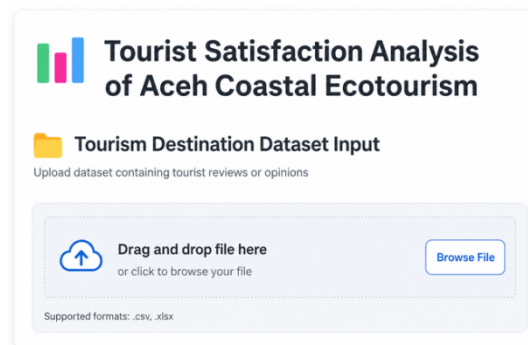


Figure 4. GUI for Dataset Upload

Figure 4 shows the initial interface of the developed GUI. The interface provides a dataset upload feature that allows users to input tourist opinion data from each tourism destination. The uploaded file is expected to contain tourist review or opinion data that will be processed by the classification model. This feature is important because it enables users to analyze tourist satisfaction data without directly interacting with the machine learning code or preprocessing pipeline.

After the dataset is uploaded, users can run the prediction process through the Process and Prediction button. The system then processes the input data, applies the trained stacking model, and generates prediction results for each tourist opinion. This workflow shows that the GUI serves as a practical bridge between the developed machine learning model and real-world usage by tourism managers.

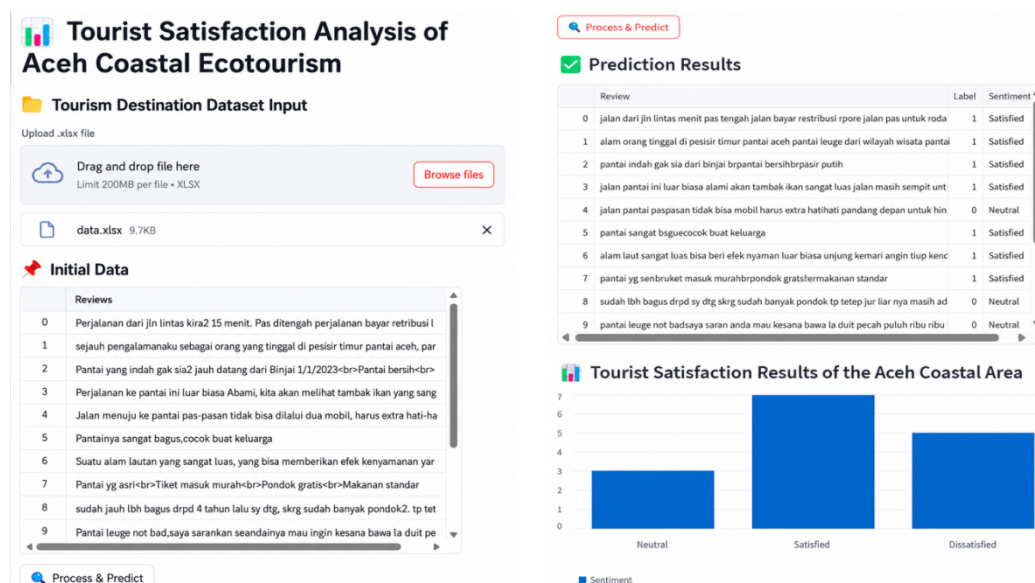


Figure 5. GUI Display of Prediction Results

Figure 5 presents the prediction results generated by the GUI. The system displays the uploaded tourist opinions along with their predicted labels and sentiment categories. Each review is classified into one of the satisfaction categories, such as Satisfied, Neutral, or Dissatisfied. This result table allows users to examine the classification output for each individual tourist opinion, making it easier to identify which reviews indicate positive, neutral, or negative satisfaction responses.

In addition to the prediction table, the GUI also presents a visualization of tourist satisfaction results in the form of a bar chart. The chart summarizes the distribution of predicted satisfaction categories from the uploaded dataset. Based on the displayed example, the system is able to show the number of tourist opinions classified as Neutral, Satisfied, and Dissatisfied. This visualization helps users understand the overall satisfaction tendency of tourists for a particular coastal tourism destination.

The GUI-based prototype provides a preliminary demonstration of how the trained stacking model can be presented through an accessible interface. While the stacking model achieved the best classification performance compared to the individual base models, the prototype is intended primarily to demonstrate the model implementation rather than to serve as a validated decision-support system. It allows users to upload tourist opinion data, review the predicted satisfaction categories, and observe the distribution of classification results. These features illustrate the potential use of the model for identifying tourist satisfaction patterns and supporting the initial interpretation of visitor opinions. However, the prototype has not yet been evaluated through usability testing, user acceptance testing, or operational deployment involving destination managers. Further validation is required before the interface can be considered suitable for practical use in coastal ecotourism service improvement.

4. CONCLUSION

Provide a statement that what is expected, as stated in the "Introduction" chapter can ultimately result in "Results and Discussion" chapter, so there is compatibility. Moreover, it can also be added the prospect of the development of research results and application prospects of further studies into the next (based on result and discussion).

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to Universitas Samudra for the support and facilities provided during the implementation of this research. This study is one of the research outputs of the Penelitian Dosen Pemula (PDP) program under contract number 625/DST/UN54.6/PT.01.03/2026. The authors also thank all parties who contributed to the completion of this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this research.

DATA AVAILABILITY

The data used in this study are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

Imam Gunanjar and Ryanda Saputra: Conceptualization; Data Curation; Methodology; Writing-Original Draft; Visualization. **Shary Armonitha Lusiana and M. Khairul Anam:** Conceptualization; Data Labelling; Methodology; Supervision; Writing-Review & Editing. **Martahadi Mardhani:** Methodology; Supervision; Writing-Review & Editing; Funding Acquisition.

REFERENCES

- [1] R. Santoso, "Tourism-Led Economic Growth in Coastal Areas: A Qualitative Study of Marine Economies," *Jurnal Ilmu Ekonomi Terapan*, vol. 10, no. 2, pp. 238–251, Dec. 2025, doi: 10.20473/jiet.v10i2.76620.
- [2] K. Hasan *et al.*, "Communication Empowerment Of Aceh East Coastal Communities In Maintaining Mangrove Ecosystems (Study Of Meunasah Asan Village, Madat District, East Aceh Regency)," in *Proceedings of International Conference on Social Science, Political Science, and Humanities (ICoSPOLHUM)*, LPPM Universitas Malikussaleh, Dec. 2022, p. 00005. doi: 10.29103/icospolhum.v3i.64.
- [3] I. Irmawati and H. Hasnawati, "Community-based Ecotourism Strategy for Local Economic Empowerment," *Journal of Education, Humaniora and Social Sciences (JEHSS)*, vol. 7, no. 2, pp. 395–404, Nov. 2024, doi: 10.34007/jehss.v7i2.2359.
- [4] R. Samal and M. Dash, "Ecotourism, biodiversity conservation and livelihoods: Understanding the convergence and divergence," *International Journal of Geoheritage and Parks*, vol. 11, no. 1, pp. 1–20, Mar. 2023, doi: 10.1016/j.ijgeop.2022.11.001.
- [5] N. P. Thang, "Understanding Eco-tourism Satisfaction: A Structural Equation Modeling Examination of Critical Determinants," *Sage Open*, vol. 15, no. 3, pp. 1–19, Sep. 2025, doi: 10.1177/21582440251378061.
- [6] Z. A. Torabi, A. A. Shalbafian, Z. Allam, Z. Ghaderi, B. Murgante, and A. R. Khavarian-Garmsir, "Enhancing Memorable Experiences, Tourist Satisfaction, and Revisit Intention through Smart Tourism Technologies," *Sustainability (Switzerland)*, vol. 14, no. 5, pp. 1–18, Mar. 2022, doi: 10.3390/su14052721.
- [7] M. L. Jin, Y. Choi, C. K. Lee, and M. S. Ahmad, "Effects of place attachment and image on revisit intention in an ecotourism destination: Using an extended model of goal-directed behavior," *Sustainability (Switzerland)*, vol. 12, no. 18, pp. 1–16, Sep. 2020, doi: 10.3390/SU12187831.
- [8] L. L. Diniz, P. M. Machado, L. L. Costa, N. B. Côgo, J. F. Madureira, and I. R. Zalmon, "Visitors' perception, coastal scenery quality and tourism profitability on sandy beaches - Southeastern Brazil," *Mar. Policy*, vol. 187, pp. 1–12, May 2026, doi: 10.1016/j.marpol.2026.107065.
- [9] P. Thakur and A. Kumar, "Service Quality and Tourist Satisfaction: A Critical Analysis of Service Delivery in Tourism Destinations," *Int. J. Innov. Sci. Res. Technol.*, vol. 11, no. 3, pp. 1009–1016, Mar. 2026, doi: 10.38124/ijisrt/26mar609.
- [10] M. K. Anam, S. Defit, and R. Sovia, "Evaluating Word2Vec and FastText Embeddings in CNN-LSTM for Hotel Customer Satisfaction: Impact on Accuracy and Model Stability," in *International Conference on Data Science and Its Applications (ICoDSA)*, Jakarta: IEEE, 2025, pp. 540–545. doi: 10.1109/ICoDSA67155.2025.11157433.
- [11] U. I. Arsyah, M. Pratiwi, H. Fryonanda, M. K. Anam, and Munawir, "ACLM Model: A CNN-LSTM and Machine Learning Approach for Analyzing Tourist Satisfaction to Improve Priority Tourism Services," *Journal of Applied Data Sciences*, vol. 6, no. 4, pp. 2891–2906, Sep. 2025, doi: 10.47738/jads.v6i4.974.
- [12] J. Ipmawati, S. Saifulloh, and K. Kusnawi, "Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 247–256, Jan. 2024, doi: 10.57152/malcom.v4i1.1066.
- [13] M. S. Hadi, J. Akbar, and M. F. Zulkarnain, "Analisis Sentimen Wisata Air Terjun Di Kabupaten Lombok Tengah Menggunakan Metode Support Vector Machine (SVM)," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 8, no. 2, pp. 318–329, 2025, doi: 10.36080/skanika.v8i2.3578.
- [14] A. Khaidar, S. Muliana, and P. Studi Magister Teknologi Informasi, "Sentiment Analysis of Instagram Comments on the BPS Province X Account Using the Naive Bayes Algorithm Based on Machine Learning," *Journal of Artificial Intelligence and Software Engineering*, vol. 5, no. 3, pp. 1231–1237, 2025, doi: 10.30811/jaise.v5i3.7815.

- [15] J. Friadi and D. E. Kurniawan, "Analisis Sentimen Ulasan Wisatawan Terhadap Alun-Alun Kota Batam: Perbandingan Kinerja Metode Naïve Bayes dan Support Vector Machine," *Jurnal Sistem Informasi Bisnis*, vol. 14, no. 4, pp. 403–407, Oct. 2024, doi: 10.21456/vol14iss4pp403-407.
- [16] D. I. Af'idah, D. Dairoh, S. F. Handayani, R. W. Pratiwi, and S. I. Sari, "Sentimen Ulasan Destinasi Wisata Pulau Bali Menggunakan Bidirectional Long Short Term Memory," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 607–618, Jul. 2022, doi: 10.30812/matrik.v21i3.1402.
- [17] M. K. Anam, T. P. Lestari, L. Efrizoni, N. S. Handayani, and I. Andhika, "Sentiment Analysis Optimization Using Ensemble of Multiple SVM Kernel Functions," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 9, no. 4, pp. 905–914, Aug. 2025, doi: 10.29207/resti.v9i4.6708.
- [18] M. K. Anam *et al.*, "Enhancing the Performance of Machine Learning Algorithm for Intent Sentiment Analysis on Village Fund Topic," *Journal of Applied Data Sciences*, vol. 6, no. 2, pp. 1102–1115, 2025, doi: 10.47738/jads.v6i2.637.
- [19] I. Setiawan *et al.*, "Utilizing Random Forest Algorithm for Sentiment Prediction Based on Twitter Data," in *Proceedings of the First Mandalika International Multi-Conference on Science and Engineering 2022, MIMSE 2022 (Informatics and Computer Science)*, Atlantis Press International BV, 2022, pp. 446–456. doi: 10.2991/978-94-6463-084-8_37.
- [20] R. Rahmaddeni, M. K. Anam, Y. Irawan, S. Susanti, and M. Jamaris, "Comparison of Support Vector Machine and XGBSVM in Analyzing Public Opinion on Covid-19 Vaccination," *ILKOM Jurnal Ilmiah*, vol. 14, no. 1, pp. 32–38, Apr. 2022, doi: 10.33096/ilkom.v14i1.1090.32-38.
- [21] Agusviyanda, Hamdani, M. K. Anam, Agustin, and M. I. Wibowo, "Stacking Ensemble Machine Learning Model for Early Detection of Chronic Kidney Disease in Indonesia," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 1244–1252, Jul. 2025, doi: 10.36341/rabit.v10i2.6501.
- [22] M. A. Latief, L. R. Nabila, W. Miftakhurrahman, S. Ma'rufatullah, and H. Tantyoko, "Handling Imbalance Data using Hybrid Sampling SMOTE-ENN in Lung Cancer Classification," *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 3, no. 1, pp. 11–18, Feb. 2024, doi: 10.30812/ijecsa.v3i1.3758.
- [23] G. Husain *et al.*, "SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models," *Algorithms*, vol. 18, no. 1, pp. 1–16, Jan. 2025, doi: 10.3390/a18010037.
- [24] S. G. Begum and P. K. Sree, "Drug Recommendation Using a 'Reviews and Sentiment Analysis' By a Recurrent Neural Network," *Indonesian Journal of Multidisciplinary Science*, vol. 2, no. 9, pp. 3085–3094, Jun. 2023, doi: 10.55324/ijoms.v2i9.530.
- [25] L. K. Ramasamy, S. Kadry, Y. Nam, and M. N. Meqdad, "Performance analysis of sentiments in Twitter dataset using SVM models," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, pp. 2275–2284, Jun. 2021, doi: 10.11591/ijece.v11i3.pp2275-2284.
- [26] V. H. Pranatawijaya, N. N. K. Sari, R. A. Rahman, E. Christian, and S. Geges, "Unveiling User Sentiment: Aspect-Based Analysis and Topic Modeling of Ride-Hailing and Google Play App Reviews," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 3, pp. 328–339, Oct. 2024, doi: 10.20473/jisebi.10.3.328-339.
- [27] S. R. Aisy and B. Prasetyo, "Sentiment Analysis of the TPKE Law on Twitter Using InSet Lexicon with Multinomial Naïve Bayes and Support Vector Machine Based on Soft Voting," *Recursive Journal of Informatics*, vol. 1, no. 2, pp. 93–101, Sep. 2023, doi: 10.15294/rji.v1i2.68324.
- [28] P. H. Prastyo, I. Ardiyanto, and R. Hidayat, "Indonesian Sentiment Analysis: An Experimental Study of Four Kernel Functions on SVM Algorithm with TF-IDF," in *International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 1–6. doi: 10.1109/ICDABI51230.2020.9325685.

- [29] F. Alzami, E. D. Udayanti, D. P. Prabowo, and R. A. Megantara, "Document Preprocessing with TF-IDF to Improve the Polarity Classification Performance of Unstructured Sentiment Analysis," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, pp. 235–242, Aug. 2020, doi: 10.22219/kinetik.v5i3.1066.
- [30] B. P. Zen, I. Susanto, and D. Finaliamartha, "TF-IDF Method and Vector Space Model Regarding the Covid-19 Vaccine on Online News," *Sinkron*, vol. 6, no. 1, pp. 69–79, Oct. 2021, doi: 10.33395/sinkron.v6i1.11179.
- [31] Z. Jiang, B. Gao, Y. He, Y. Han, P. Doyle, and Q. Zhu, "Text Classification Using Novel Term Weighting Scheme-Based Improved TF-IDF for Internet Media Reports," *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/6619088.
- [32] Herianto, B. Kurniawan, Z. H. Hartomi, Y. Irawan, and M. K. Anam, "Machine Learning Algorithm Optimization using Stacking Technique for Graduation Prediction," *Journal of Applied Data Sciences*, vol. 5, no. 3, pp. 1272–1285, Sep. 2024, doi: 10.47738/jads.v5i3.316.
- [33] M. K. Anam, T. Nasution, H. Yenni, M. Pratiwi, U. I. Arsyah, and I. Syahputra, "Improving the Performance of Naïve Bayes Using Stacking Technique for Detecting Stress in Students in Indonesia," in *2025 2nd International Conference on Information System and Information Technology (ICISIT)*, IEEE, Nov. 2025, pp. 1–6. doi: 10.1109/ICISIT66233.2025.11402979.