BIMA
BULLETIN OF INTELLIGENT MACHINES AND ALGORITHMS
Vol. 1 No.4 May 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i4.12

A Comparative Study of Machine Learning Models for Airbnb Booking Likelihood Prediction in Singapore with GridSearchCV Optimization

Rizal Habibulloh[1*], Rafi Mohammad Alhafidz[2], Agung Prayitno[3]

Informatics[123]
Universitas Informatika dan Bisnis Indonesia[123]
https://unibi.ac.id/[123]
rizal.h22@student.unibi.ac.id[1*], rafi.ma22@student.unibi.ac.id[2], agung.p22@student.unibi.ac.id[3]

Abstract. The rapid development of digitization has driven a fundamental shift in the travel industry, particularly through the emergence of shared economy platforms like Airbnb. Thus, the requirement to have a prediction model for travel industry is really crucial point and will give huge benefit to travel industry to solve their problem to finding host to rent their building for the industry and guest to rent for property rented from the platform. From this study we have several prediction models which developed from 10 machine learning algorithm. Each model has a distinct of its own accuracy f1 score and ROC AUC which each score would reveal how good is each model to be utilized for booking likelihood prediction. Some Algorithm like XGboost, Random Forest, Logistic Regression, SVM has huge of accuracy for each score to extend from 90% accuracy after we did a hyperparameter tunning to boost each model's performances. However, these three score (accuracy, f1 score, and roc auc) aren't sufficient to make the model work efficient and to be reliable to be utilized for our prediction model. Hence, more analytical methods are required to make sure our models are perform well for our aim to create a reliable prediction model and less bias on output we desired. In this study we shall commit curve learning analysist to make our models surely become a more reliable models which give numerous benefits to model accuracy and which many studies are still overturned the methods that very unfortunate. We shall dismantle our discovery and reveal which machine learning algorithm, optimization technique and how to analyst learning curve for each model on its own chapter hopefully our research is useful for everyone to gain new knowledge to developing prediction model using machine learning algorithm.

Keywords: machine learning; booking likelihood prediction; optimization; GridSearchCV; SVM

1. INTRODUCTION

The rapid development of digitization has driven a fundamental shift in the travel industry, particularly through the emergence of shared economy platforms like Airbnb, now one of the largest accommodation booking providers globally[1]. The platform offers a vast number of listings with a broad range of characteristics, including price, location, amenities, user review quality, and cancellation policy flexibility. This level of diversity creates a highly competitive market environment, requiring property owners (hosts) to have a deep understanding of the factors that affect the likelihood of a property being selected and booked by potential guests[2]. In this context, the capacity to estimate booking likelihood or probability of booking is a crucial component in supporting data-driven decision-making for hosts, platform managers, and other stakeholders.

Significance of booking likelihood prediction issues is increasingly growing along with the increasing amount of available properties and dynamic changes in consumer preferences and behavior. Platform users typically evaluate various alternatives before deciding to book, so the likelihood of a property getting booked is no longer determined by a single factor, but by a complex interaction of various attributes[3]. An inability to identify the key determinants that influence booking decisions has the potential to result in ineffective pricing strategies, suboptimal occupancy rates, and economic losses for hosts[4]. Furthermore, the platform's perspective, inaccuracies in predicting booking likelihood may also reduce the quality of the recommendation system and may negatively impact the overall user experience.

Numerous studies have examined the impact of pricing, customer reviews, and user behavior patterns upon the Airbnb ecosystem[5]. Despite this, the challenge of

Vol. 1 No.4 May 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i4.12

accurately predicting booking likelihood remains largely unresolved. This is due to the complex, heterogeneous, high-dimensional nature of Airbnb data, which also contains nonlinear relationships between variables[6]. In addition, the unbalanced distribution of data, where the number of listings that are actually booked is relatively smaller than those that are not booked in a given period, poses a particular challenge in the modeling process. This complexity is compounds by the influence of external factors such as seasonality, tourism trends, and local user preferences, thereby making it difficult to build a stable prediction model with good generalization capabilities. Given these conditions, traditional statistical approaches are often inadequate for capturing the latent patterns hidden within the data.

The significant progress in the fields of Machine Learning and Data Science presents a strategic opportunity to tackle these challenges. Machine learning methodologies can represent complex nonlinear relationships, manage high-dimensional data, and extract important patterns that are challenging to identify through conventional analysis[7]. Nevertheless, the implementation of Machine Learning in predicting booking likelihood is focused not only on achieving high accuracy but also on ensuring adequate interpretability[8]. These aspects are important so that the modeling results can be comprehended and practically utilized by non-technical users, particularly hosts and accommodation business managers. Consequently, an approach that balances the predictive performance of the model with a clear understanding of the factors determining the prediction results is required. Based on this literature, this study seeks to design and evaluate a Machine Learning model capable of predicting booking likelihood on the Airbnb platform by capitalizing on listing features and user review information. The main focus of the research is oriented towards identifying features that have a significant influence on booking likelihood, alongside a comparative analysis of the performance of various Machine Learning algorithms in handling the complexity of Airbnb data. Furthermore, this research also seeks to explore customer booking behavior patterns through the interpretation of the modeling results produced.

This research is intended to have a significant impact both theoretically and practically. Academically, this study enriches the literature on the application of machine learning in the context of digital tourism, particularly in the area of predicting booking probabilities. Practical implications, The findings of this study may be utilized by Airbnb hosts as a basis for formulating pricing strategies, enhancing service quality, and optimizing listing performance to strengthen competitiveness. Moreover, the research outcomes may also support the development of recommendation systems and data-driven decision-making mechanisms on the Airbnb platform in a more effective and sustainable manner.

2. RELATED WORK

Previous research is a review of earlier studies that are related to the topic of the current research. The purpose of this literature review is to identify the development of research, methodological approaches that have been employed, as well as the differences and novelty of the proposed research[9]. A number of previous studies have discussed the application of machine learning in Airbnb booking classification, notably the journal article entitled Airbnb listings' performance: determinants and predictive models by Efstathios Kirkos[10]. This journal article shows that preliminary data analysis reveals interesting aspects of the Airbnb market in Thessaloniki. The other interesting issue is that a relatively low proportion (19%) of hosts have verified their identities. Studies show that identity verification increases perceived quality among potential customers[11]. Based on this study, we can conclude that machine learning is effective in assisting the decision-making process. However, there are still some differences in variables, criteria, and research objects. For this reason, this study was conducted to use a system of several machine learning algorithms to find out which algorithm is the most effective.

Some other study that given common goal with our research is a study named "Hotel Booking Prediction using Machine Learning" by Pranav Kumar [12]. He revealed he and co has choice the model that common like this research for example Logistic Regression, KNN, Decision Tree, Random Forest and XGboost. They got 80% average for their model accuracy and got a higher value after doing a parameter tunning which raised the accuracy value up to 85%.In other study Efraim William Solang also have common interest to testing machine learning model algorithms to be utilized for creating model prediction using a classification algorithm in one of his article called "Machine Learning Evaluation for Hotel Cancellation Prediction with Threshold Adjustment and Cost-Based Evaluation"[13] even though the goal is quite distinct from our research but still using the same method also the very common algorithm.

3. METHODS

This research employs a quantitative approach based on data analysis incorporating machine learning techniques to model occupancy rates or booking likelihood on the Airbnb platform in Singapore. A structured and sequential methodological framework was designed, covering data collection, exploratory analysis, data preprocessing, algorithm selection, model training and validation, hyperparameter optimization, and comprehensively evaluating the performance of the developed model. Here's the pipeline that we have implemented during this study as shown in Figure 1.

BIMA
BULLETIN OF INTELLIGENT MACHINES AND ALGORITHMS
Vol. 1 No.4 May 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i4.12

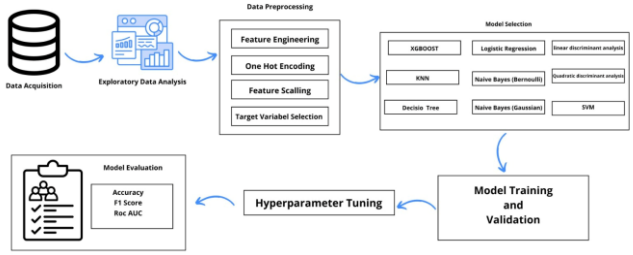

Fig.1. Study's Pipeline Process

3.1. Data Acquisition

Data Acquisition as shown in Table 1 is an approach to choosing a dataset to be utilized for machine learning model training [14].This study employs Airbnb Singapore's dataset processed in Python programming environment by using Google Collab. The dataset holds a number of variables representing property attributes and booking activity indicators. We define the target variable as booked in binary form in order to represent the presence of a property booking indication, derived from availability information and review activity.

TABLE 1. Dataset's Feature

Feature Name	Description
price	Property rental price per night stated in local currency.
minimum_nights	The minimum number of nights required to make a reservation at a property.
availability_365	The number of days in a year (365 days) during which the property is available for booking.
number_of_reviews	The total number of reviews received by a property since it was listed on the platform.
reviews_per_month	The total number of reviews received by a property since it was listed on the platform.
calculated_host_listings_count	The total number of properties managed by a single host as an indicator of ownership scale.
room_type	The types of accommodation offered, such as entire home, private room, or shared room.
neighbourhood_group	The administrative region or group of areas where the property is located in Singapore.
booked	A binary target variable representing property booking status (1 = booked, 0 = not booked)..

3.2. Exploratory Data Analysis (EDA)

Exploratory data analysis is performed to obtain an initial overview of the distribution and characteristics of each variable in the dataset[15]. Descriptive statistics apply to numerical features to recognize data distribution patterns and the presence of extreme values. The reviews_per_month feature has been analyzed in more depth given the high proportion of missing values. Moreover, distribution analysis has been performed on categorical variables to comprehend the variation in accommodation types and location distribution. The class distribution in the target variable booked has been evaluated to identify potential class imbalances that could influence the classification model's performance.

3.3. Data Preprocessing.

The preprocessing stage is initiated by handling missing values, notably in specific variables, which are imputed with a value of zero to indicate the absence of review activity. Categorical variables will then be transformed using the one-hot encoding technique to make them compliant with the Machine Learning algorithm. Subsequently, all numerical features will be normalized to ensure consistency in scale between variables. Once the entire transformation process is complete, the dataset will be separated into a feature matrix (X) and target variable (y), with booked as the classification label. In this study, the dependent variable (Y) is will_rebook, which is defined as a binary variable with a value of 0 or 1. This variable indicates the likelihood of rebooking a property, which is determined based on the existence of reviews in the last 12 months. Operatively, a property is categorized as having a value of will_rebook = 1 if the number of reviews in the last 12 months (number_of_reviews_ltm) is greater than zero, and a value of 0 if the opposite is true.

Meanwhile, the independent variables (X) encompass all attributes contained in df_processed_new, with the exception of the target variable will_rebook. This set of feature variables comprises a combination of numerical and categorical features that have undergone data preprocessing. Numeric features were normalized or scaled to guarantee the stability and consistency of the model learning process, including price, minimum_nights, number_of_reviews,reviews_per_month,calculated_host_listings_count, availability_365, price_per_night, and review_engagement.

3.4. Model Selection

This study evaluates ten supervised learning-based classification algorithms, namely Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree Classifier, Random Forest Classifier, XGBoost Classifier, Gaussian Naive Bayes, Bernoulli Naive Bayes, Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA).

3.5. Model Training and Validation

The dataset is separated into training data and test data using a train-test split scheme. We trained each model using training data and evaluated it through cross-validation to ensure the stability of the model's performance. This approach aims to ensure that the model not only performs well on training data but also has adequate generalization capabilities on data that has yet

Vol. 1 No.4 May 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i4.12

to be used in the training process. Furthermore, we visualized learning curves to analyze indications of overfitting and underfitting in each model.

3.6. Hyperparameter Tuning

Hyperparameter optimization using the GridSearchCV approach was performed on each classification algorithm tested. These steps aimed to identify the combination of hyperparameters leading to the best performance based on cross-validation evaluation. The optimal hyperparameters obtained were then used to retrain the models on the training data, so that each model was evaluated in its most representative configuration before the final comparison. The redefinition of the target variable to target variabel (will_rebook) has been completed as the presence of reviews in the last twelve months (number_of_reviews_ltm > 0) which represents as considered effort to mitigate data leakage inherent in previous formulations. Nevertheless, this proxy for booking likelihood is risk from several limitations for example it remains an indirect and lagged indicator of actual booking volume, is susceptible to biases stemming from guest review submission behaviors, and may not fully capture the continuous nature of occupancy dynamics. Acknowledging these constraints is crucial for a nuanced interpretation of model performance and for guiding future research toward more direct measures of booking success, potentially through regression-based approaches or more granular event data. To ensure methodological rigor and enhance model performance, a comprehensive hyperparameter optimization was executed for each classification algorithm via GridSearchCV employing 5-fold cross-validation. This systematic search explored predefined parameter spaces to identify optimal configurations. Here is detailed parameter we have applied to improve our models' performances:

Logistic Regression

In the Logistic Regression model, the hyperparameter tuning process focused on the values of C and the type of regularization penalty. The parameter C was tested with values of 0.1, 1, and 10 to control the inverse strength of regularization. Smaller values of C apply stronger regularization, which helps prevent overfitting by penalizing large coefficient values and encouraging simpler models. Conversely, larger values allow the model to fit the training data more closely, potentially increasing model complexity. Therefore, tuning the C parameter is important to achieve a balance between model simplicity and predictive performance. In addition, the model evaluated both L1 and L2 regularization penalties. L1 regularization promotes sparsity by driving some coefficients to zero, thereby performing implicit feature selection and improving interpretability. On the other hand, L2 regularization reduces coefficient magnitudes without eliminating them completely, which helps minimize multicollinearity and improve model stability. Selecting the appropriate penalty type can significantly enhance the model's generalization capability.

Decision Tree

For the Decision Tree model, several hyperparameters were optimized, including max_depth, min_samples_split, and min_samples_leaf. The max_depth parameter was evaluated using values of 2, 5, 10, and None to control the maximum depth of the tree. Limiting the tree depth reduces model complexity and helps prevent overfitting, whereas deeper trees are capable of capturing more detailed patterns in the training data but may generalize poorly. The value None allows the tree to grow until all leaves are pure or contain fewer samples than the minimum split threshold. The min_samples_split parameter, tested with values of 2 and 5, determines the minimum number of samples required to split an internal node. Increasing this value prevents the model from learning overly specific patterns from small subsets of data, thereby improving robustness. Additionally, the min_samples_leaf parameter, tested with values of 1 and 2, controls the minimum number of samples required at a leaf node. This parameter helps smooth the model and reduces overfitting by ensuring that leaf nodes are not formed from very small sample groups.

Random Forest

In the Random Forest model, hyperparameter tuning was performed on n_estimators, max_depth, min_samples_split, and min_samples_leaf. The n_estimators parameter, evaluated with values of 50 and 100, determines the number of trees in the forest. Increasing the number of trees generally improves model performance by reducing variance through averaging multiple predictions, although excessive numbers of trees may increase computational cost without substantial performance gains. The max_depth parameter, tested with values of 2, 5, and None, controls the depth of each individual tree. Limiting tree depth helps reduce overfitting and contributes to better ensemble generalization. The min_samples_split parameter, evaluated using values of 2 and 5, specifies the minimum number of samples required to split an internal node, thereby preventing individual trees from becoming overly specialized. Similarly, the min_samples_leaf parameter, tested with values of 1 and 2, ensures that leaf nodes contain a minimum number of samples, improving the stability and robustness of the forest model.

Support Vector Machine (SVM)

For the Support Vector Machine (SVM) model, the hyperparameters C, kernel, and gamma were optimized. The C parameter was evaluated with values of 1 and 10 to control the balance between maximizing the decision boundary margin and minimizing classification errors. Smaller values of C apply stronger regularization and

Vol. 1 No.4 May 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i4.12

encourage wider margins, which may improve generalization but allow more training errors. In contrast, larger values prioritize correct classification of training samples, potentially increasing the risk of overfitting. The model utilized the Radial Basis Function (RBF) kernel, which enables SVM to model non-linear relationships by projecting data into higher-dimensional feature spaces. Furthermore, the gamma parameter was set to scale, which automatically computes the gamma value based on the number of features and data variance. This setting often provides stable and robust generalization performance by controlling the influence range of individual training samples on the decision boundary.

K-Nearest Neighbors (KNN)

In the K-Nearest Neighbors (KNN) model, the hyperparameters n_neighbors and weights were tuned to optimize classification performance. The n_neighbors parameter was tested using values of 3, 5, and 7 to determine the number of neighboring samples considered during classification. Smaller values of k increase model sensitivity to local patterns and noise, which may lead to overfitting, while larger values provide smoother decision boundaries but may overlook important local structures. Therefore, selecting an appropriate number of neighbors is essential for balancing bias and variance. The weights parameter was evaluated using uniform and distance. Uniform weighting assigns equal importance to all neighboring samples, whereas distance weighting gives greater influence to closer neighbors. The distance-based approach can improve model sensitivity to nearby data points and potentially enhance classification accuracy.

XGBoost (Extreme Gradient Boosting)

For the XGBoost model, several important hyperparameters were optimized, including n_estimators, max_depth, learning_rate, subsample, and colsample_bytree. The n_estimators parameter, tested with values of 50 and 100, controls the number of boosting iterations or trees in the ensemble. Increasing this value generally improves model performance, although excessive boosting rounds may increase computational complexity and the risk of overfitting. The max_depth parameter, evaluated with values of 3 and 5, determines the complexity of individual trees, where smaller values help regularize the model and improve generalization. The learning_rate parameter, tested with values of 0.01 and 0.1, controls the contribution of each boosting step by shrinking feature weights after every iteration. Smaller learning rates make the training process more conservative and often require more trees to achieve optimal performance. In addition, subsample and colsample_bytree were evaluated using values of 0.7 and 1.0. These parameters randomly sample training instances and features, respectively, during tree construction, thereby reducing variance, increasing model diversity, and preventing overfitting.

Naive Bayes (Gaussian)

In the Gaussian Naive Bayes model, the hyperparameter var_smoothing was optimized using values of 1e-9, 1e-5, and 1e-1. This parameter adds a small portion of the largest feature variance to all variances to improve numerical stability during probability calculations. The smoothing process prevents zero variance issues that may lead to unstable computations, such as division-by-zero errors. As a result, var_smoothing acts as a regularization mechanism that stabilizes likelihood estimation and improves model robustness, particularly when dealing with noisy or limited datasets.

Naive Bayes (Bernoulli)

For the Bernoulli Naive Bayes model, the alpha parameter was optimized using values of 0.01, 0.1, and 1.0. This parameter represents additive smoothing, also known as Laplace or Lidstone smoothing, which prevents zero probabilities when a feature value is absent in a particular class during training. Smaller alpha values apply minimal smoothing, whereas larger values increase smoothing effects and improve the model's ability to handle unseen feature combinations. Proper tuning of alpha is important for enhancing classification stability and improving generalization performance.

Linear Discriminant Analysis (LDA)

In the Linear Discriminant Analysis (LDA) model, the hyperparameters solver and shrinkage were evaluated. The solver parameter was tested using svd and lsqr. The svd solver is considered robust because it does not require explicit covariance matrix computation and can handle singular covariance matrices effectively. Meanwhile, the lsqr solver supports shrinkage regularization, making it suitable for high-dimensional or highly correlated datasets. The shrinkage parameter, applied only with the lsqr solver, was tested using values of None, 0.5, and 0.9. Shrinkage regularization stabilizes covariance matrix estimation by pulling it toward a scaled identity matrix, which improves numerical stability and model generalization, especially when the number of features is large relative to the number of training samples.

Quadratic Discriminant Analysis (QDA)

For the Quadratic Discriminant Analysis (QDA) model, the reg_param hyperparameter was optimized using values of 0.0 and 0.1. This parameter introduces regularization into the covariance matrix estimation process, helping prevent singular or ill-conditioned covariance matrices, particularly when working with small datasets or highly correlated features. A value of 0.0 indicates no regularization, whereas higher values apply stronger regularization, improving numerical stability and enhancing the model's ability to generalize to unseen data.

Vol. 1 No.4 May 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i4.12

3.7. Model Evaluation

Model performance evaluation is performed using test data with reference to three main metrics, namely accuracy, F1-score, and ROC AUC. Evaluations of all optimized models are systematically compared to determine the model with the best performance in predicting booking likelihood.

4. RESULTS AND DISCUSSIONS

4.1. RESULTS

Once the target variables have been redefined and the hyperparameter tuning process has complete using GridSearchCV across all ten classification models, the evaluation results has shown significant improvements compared to the previous experiment. Specifically, the perfect accuracy (1.00) which showed overfitting previously achieved by several models no longer appeared. This indicates that the excessively high performance in the initial experiment was likely influenced by data leakage or an overly simplistic definition of target variables, which did not realistically represent the complexity of the problem. Therefore, the latest evaluation results are assessed to be more valid and reflect the actual generalization capabilities of the model.

The cross-validation results indicate that the XGBoost model remains the approach with the best performance, even though its accuracy value has dropped from previous results. However, this decrease actually strengthens the model's credibility because it reflects a more rigorous learning process that is free from structural bias in the data. The Decision Tree and Random Forest models also show promising performance with relatively small accuracy differences compared to XGBoost, indicating that tree-based methods are still effective in capturing nonlinear patterns in Airbnb data. In addition, several models showed improvement on their performance after target redefinition, particularly Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Bernoulli Naive Bayes, and Linear Discriminant Analysis (LDA). This improvement demonstrates that these models are better able to accommodate target formulations that are more representative of the reordering phenomenon. Thus, Gaussian Naive Bayes and Quadratic Discriminant Analysis (QDA) experienced a significant decline in accuracy, indicating high sensitivity to data distribution assumptions and changes in target definitions.

Overall, this assessment confirms that redefining the target variable not only reduces the risk of overfitting and data leakage but also provides a more balanced view of the strengths and limitations of each algorithm. A comparison of model accuracy after hyperparameter tuning is summarized in Table 2.

TABLE 2. Model Result After Optimization Technique

Classification	Accuracy (Before)	Accuracy (After)	F1-Score (Before)	F1-Score (After)	ROC AUC (Before)	ROC AUC (After)
Logistic Regression	1.0000	0.9508	1.0000	0.8626	1.0000	0.9548
Decision Tree	1.0000	0.9536	1.0000	0.8803	1.0000	0.9710
Random Forest	1.0000	0.9577	1.0000	0.8864	1.0000	0.9898
Support Vector Machine (SVM)	0.6790	0.9481	0.6648	0.8527	0.7358	0.9646
K-Nearest Neighbors (KNN)	0.7746	0.9385	0.7563	0.8193	0.8447	0.9502
XGBoost	1.0000	0.9617	1.0000	0.9007	1.0000	0.9904
Gaussian Naive Bayes	0.9781	0.8402	0.9783	0.5145	0.9785	0.8090
Bernoulli Naive Bayes	1.0000	0.8634	1.0000	0.6599	1.0000	0.8914
Linear Discriminant Analysis (LDA)	0.6954	0.8620	0.6421	0.6529	0.7292	0.8948
Quadratic Discriminant Analysis (QDA)	0.5260	0.8661	0.6705	0.5000	0.5332	0.8902

At first time we perform our model to train the dataset we have achieved reasonable result, some model can get various of accuracy for example Gausian Naive Bayes has achieved the best model to perform with 0.9781 accuracy, 0.9783 F1 Score, and ROC AUC 0.9785 considered to achieved highest model performance and fit model performance. Some other models have achieved result perfect accuracy for example Logistic Regression, Decision Tree, Random Forest, XGBoost, Bernoulli Naive Bayes, however this perfection doesn't show a good characteristic fit model performance. Otherwise, it's usually led to bias and risk model lead to overfitting. K-Nearest Neighbor is another model to get fit in term of accuracy, it has 0.7466 accuracy score, F1 Score 0.7563 and ROC AUC 0.8447 shown the model balanced perform to our dataset. SVM and Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) has shown quite low with bellow 70% of accuracy. However, after we run optimalization using GridsearchCV we can conclude that every model we have utilized for this research is fit with high accuracy, although some model has smaller accuracy level and also some model has f1 score quite lower to 0.51 that not a good choice for our model. However, to rely only on accuracy, f1 and roc auc score only is not really show fully the model is impressive. We need to furthermore made addition steps to ensuring that our model is still not

Vol. 1 No.4 May 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i4.12

overfitting or underfitting. The next step to create more valid analysist is to committing learning curve analysist like we shown in right side.

Figure 2 shows the learning curves of ten classification models, which represent the relationship between the quantity of training data and the model's performance on training and cross-validation data. This analysis seeks to evaluate the model's generalization ability, bias–variance trade-off level, and performance stability as the number of training samples increases.

In the Logistic Regression model, there is a relatively consistent gap between the training and cross-validation scores. Despite the training score tend to stabilize after a specified amount of data, the validation score increases gradually but does not fully converge with the training curve. This pattern indicates the limitations of linear models in capturing the nonlinear relationships present in the data, resulting in moderate underfitting despite the increased amount of data.

Tree-based models, namely Decision Tree and Random Forest, exhibit different characteristics. Decision Tree shows consistently with high training scores from the outset, with the gap to validation scores narrowing as more training data is added even leave quite a gap. This indicates that adding data helps reduce model variance. On the other hand, Random Forest demonstrates training scores across the entire data range, paired with a steady increase in validation scores but still missing many consistencies and most huge gap. This pattern shows that ensemble learning still lead to overfitting, although indications high score accuracy but not being supported by its curve which still imbalanced.

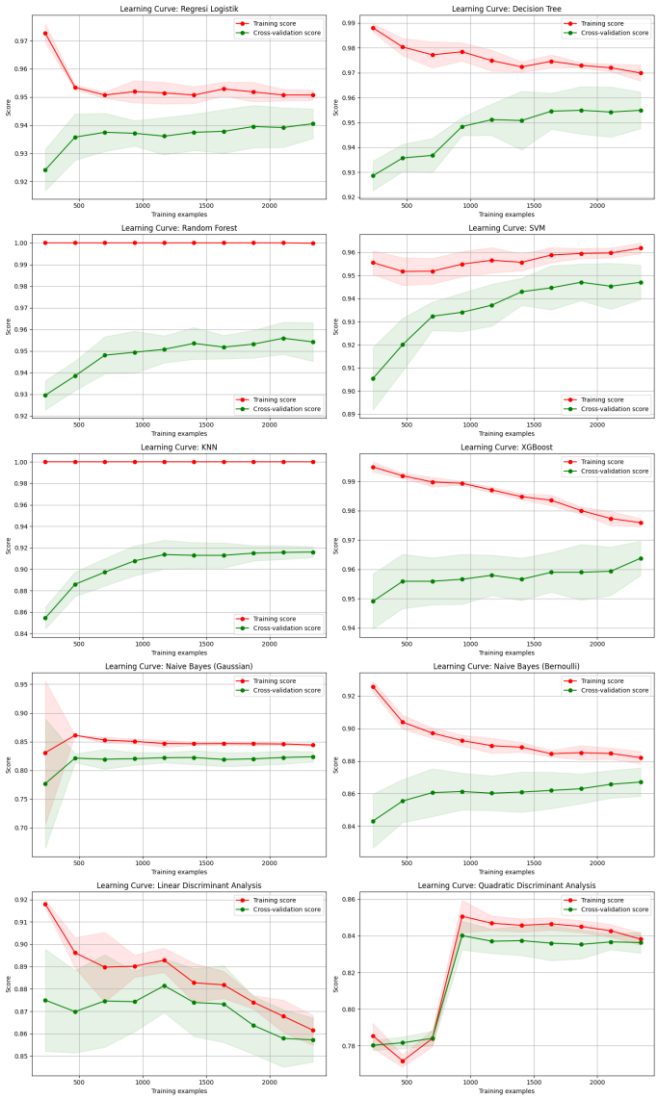

Fig.2. Learning Curve of All Model After Optimization

In Support Vector Machine (SVM), the training curve is relatively stabilized with a slight upward trend, whereas the validation curve shows a significant increase in the early phase before reaching a plateau. The gradual convergence between the two curves indicates that SVM is able to utilize additional data to improve generalization, with a relatively good balance of bias and variance. Although at first given quite a gap but SVM model gave most consistent curve with highest accuracy value.

The K-Nearest Neighbors (KNN) model demonstrates very high and consistent training scores, while validation scores rose slowly but remained at a lower level. This pattern reflects a high variance tendency, where the model is highly sensitive to training data and requires a large amount of data to achieve more stable generalization. However, the training score perfection and

Vol. 1 No.4 May 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i4.12

huge gap between two curves show this model still work imperfect.

The XGBoost model demonstrates balanced with minimum 500 dataset learning curve behavior. The training score decreases slightly as the data increases, while the validation score increases and remains relatively stable. The decreasing distance between the two curves suggests strong generalization capabilities however still leave gap which could lead to overfitting.

In the probabilistic model, Gaussian Naive Bayes demonstrates significant fluctuations with a small amount of training data, accompanied by rapid convergence in training and validation scores. However, the relatively low performance level reflects the limitations of the normal distribution assumption regarding data complexity. In contrary, Bernoulli Naive Bayes demonstrated a more stable curve with a gradual increase in validation scores, indicating a better fit to the characteristics of binary features resulting from one-hot encoding. Thus, the two probabilistic model is fit for this prediction model but can't be regarded as the best model with less accuracy than 90%.

The Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) models demonstrated fairly high sensitivity to data size. LDA demonstrates a downward trend in validation scores with large amounts of data, indicating the limitations of linear assumptions in class separation. QDA demonstrates a sharp increase in performance after a certain data threshold, but still displays a gap between training and validation scores, reflecting sensitivity to covariance estimation in high-dimensional data. Overall, this learning curve analysis demonstrates that ensemble models, even with highest score XGBoost primarily indicate a gap and potentially gave more bias occurred, thus even it's able to possess the higher score compared to other models. These results justify the selection of XGBoost cannot be judge as the primary model in this study, as it exhibits a quite imbalance between high accuracy, stability with respect to data additions, and risk of overfitting, instead SVM is the most consistency model for this likelihood prediction model. Here is confusion matrix we can compare between SVM and XGBoost:

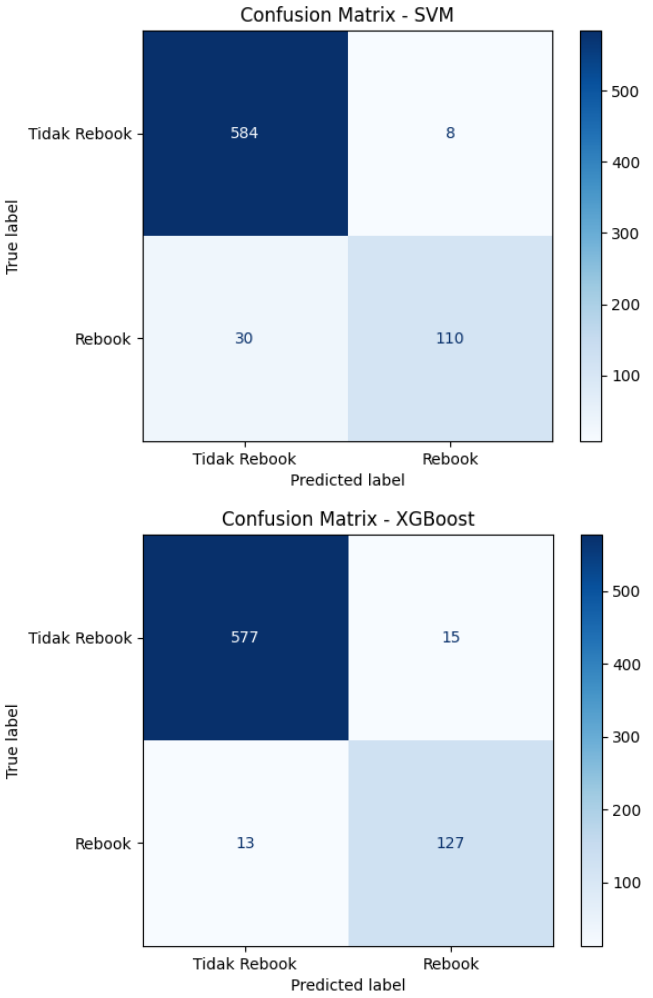

Fig.3. XGBoost and SVM Confusion Matrix

Both models have got very similar result which indicate good consistency and show that these two models are fit for classification model, as shown in Figure 3. SVM concluding that true positive 110 client will reconsider to rebooking the property whereas XGBoost predicting 127 clients. This discovery has shown the very small gap between two model and that the two models is justified one another prediction. The false positive meaning when the model is wrong to predicted that client will not rebook but in reality, is rebooking is for both 30 and 13 still show that only fewer than the correct prediction is.

4.2. DISCUSSIONS

The results of experimental has demonstrate that our model have achieved impressive and balanced performance across all models, with consistently high accuracy, recall, and F1-scores. The classification likelihood prediction report has sufficed to demonstrate each model's robustness to gained accurate booking prediction. We have achieved to get more robust from several previous study with advanced steps to using hyperparameter tunning and further with learning curve analysis.

Vol. 1 No.4 May 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i4.12

When viewed in the context of previous studies on Airbnb and hotel booking predictions, the results of this study strengthen previous findings that machine learning techniques are a powerful tool for supporting data-driven decision making in the accommodation market. Previous studies have largely focused on the identification of influential variables, descriptive market characteristics, or comparisons of single model performance. In comparison, this study amplifies this approach by systematically examining various classification algorithms within an integrated experimental framework, facilitating a more objective comparison of model effectiveness in predicting booking likelihood.

Moreover, although previous studies have demonstrated performance improvements through parameter optimization, these studies frequently rely on limited evaluation criteria and do not explicitly assess learning behavior or generalization stability. This study addresses this gap by integrating hyperparameter tuning with learning curve analysis, yielding a more in-depth examination of how different algorithms respond to increased training data and complex feature interactions.

The results demonstrate that ensemble-based models, particularly XGBoost and Random Forest, exhibit excellent robustness and consistency, supporting their selection as effective predictive tools for estimating booking likelihood. These results highlight the methodological novelty of the proposed approach and confirm its contribution beyond the existing literature in short-term rental and hospitality analysis. Here is model comparison result from two previous study as shown in Table 3.

TABLE 3. Comparison Table

Study	Classification Model	Accuracy	F1-Score	ROC AUC
[10]	Logistic Regression	0.7568	0.773	0.842
	Decision Tree (C4.5)	0.7680	0.772	0.798
	Random Forest	0.8262	0.826	0.915
	Support Vector Machine (SMO)	0.7439	0.760	0.743
[12]	Logistic Regression	0.8040	0.6989	–
	Decision Tree	0.8379	0.7722	–
	Random Forest	0.8537	0.7937	–
	K-Nearest Neighbors (KNN)	0.8442	0.7801	–
	XGBoost	0.8505	0.7887	–
This Study	Logistic Regression	0.9508	0.8626	0.9548
	Decision Tree	0.9536	0.8803	0.9710
	Random Forest	0.9577	0.8864	0.9898
	Support Vector Machine (SVM)	0.9481	0.8527	0.9646
	K-Nearest Neighbors (KNN)	0.9385	0.8193	0.9502
	XGBoost	0.9617	0.9007	0.9904

Based on the table we can conclude that we have overcome of both study's model performance and also based on Table 2 we have even more insight to discover knowledge about some other algorithm we can utilized for classification model in particular the statistical model algorithm like QDA and LDA which they also work pretty well.

5. CONCLUSION

This study evaluates the performance of various machine learning algorithms in predicting order likelihood using a rigorous evaluation framework that combines performance metrics and learning behavior analysis. Experimental results indicate that while some models attain high accuracy, F1 scores, and ROC AUC values, these metrics alone are unable to fully reflect model reliability. Notably, ensemble-based methods such as XGBoost and Random Forest yielded the highest prediction scores; however, learning curve analysis revealed persistent gaps between training and validation performance, indicating potential bias-variance imbalance and higher overfitting risk.

By incorporating learning curve diagnostics alongside conventional evaluation metrics, this study provides a more robust appraisal of the model's generalization capabilities. Analysis demonstrates that Support Vector Machine (SVM), despite achieving slightly lower peak accuracy than XGBoost, exhibits the most consistent learning behavior, with more stable convergence between training and validation curves as data volume increases. This consistency suggests a better balance between bias and variance, making SVM a more reliable choice for predicting order likelihood in practical application scenarios where robustness and stability are critical.

In contrast to previous studies that largely focused on identifying influential variables or improving accuracy through parameter adjustment, this study provides additional methodological value by highlighting diagnostic evaluation and comparative learning behavior between models. These insights highlight that model selection should not be based solely on numerical performance scores, but also on generalization stability and resilience to overfitting. Ultimately, this study provides a more holistic model evaluation strategy for predicting booking likelihood, which enriches the existing literature and offers practical insights for researchers and practitioners seeking reliable decision support systems on short-term rental platforms.

REFERENCE

Vol. 1 No.4 May 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i4.12

[1] E. A. Ndaguba and C. Van Zyl, 'Exploring bibliometric evidence of Airbnb's influence on urban destinations: emotional solidarity, Airbnb supply, moral economy and digital future', *Int. J. Tour. Cities*, vol. 9, no. 4, pp. 894–922, Nov. 2023, doi: 10.1108/IJTC-03-2023-0056.

[2] M. Rossi, 'Competition and Reputation in an Online Marketplace: Evidence from Airbnb', *Manag. Sci.*, vol. 70, no. 3, pp. 1357–1373, Mar. 2024, doi: 10.1287/mnsc.2023.4758.

[3] X. Yang and M. Tian, 'To see and then to believe: how image affect tenant decision-making and satisfaction on short-term rental platform', *Electron. Commer. Res.*, vol. 24, no. 4, pp. 2877–2901, Dec. 2024, doi: 10.1007/s10660-022-09622-z.

[4] E. Zwalnan, 'INVESTIGATING THE IMPACT OF DYNAMIC PRICING STRATEGIES ON REVENUE OPTIMIZATION IN THE', *Int. J.*, vol. 05, no. 7, 2024.

[5] M. He, 'Customer engagement and bias in online reviews: comparative analysis between Airbnb and traditional hotels', 2025, doi: 10.5525/GLA.THESIS.85658.

[6] L. Di Persio and E. Lalmi, 'Maximizing Profitability and Occupancy: An Optimal Pricing Strategy for Airbnb Hosts Using Regression Techniques and Natural Language Processing', *J. Risk Financ. Manag.*, vol. 17, no. 9, p. 414, Sep. 2024, doi: 10.3390/jrfm17090414.

[7] Nitin Liladhar Rane, Mallikarjuna Paramesha, Saurabh P. Choudhary, and Jayesh Rane, 'Machine Learning and Deep Learning for Big Data Analytics: A Review of Methods and Applications', Jun. 2024, doi: 10.5281/ZENODO.12271006.

[8] I. Gómez-Talal, M. Azizsoltani, P. Talón-Ballestero, and A. Singh, 'Machine Learning in Hospitality: Interpretable Forecasting of Booking Cancellations', *IEEE Access*, vol. 13, pp. 26622–26638, 2025, doi: 10.1109/ACCESS.2025.3536094.

[9] A. B. Nassif, M. A. Talib, Q. Nasir, Y. Afadar, and O. Elgendy, 'Breast cancer detection using artificial intelligence techniques: A systematic literature review', *Artif. Intell. Med.*, vol. 127, p. 102276, May 2022, doi: 10.1016/j.artmed.2022.102276.

[10] E. Kirkos, 'Airbnb listings' performance: determinants and predictive models', *Eur. J. Tour. Res.*, vol. 30, p. 3012, 2022, doi: 10.54055/ejtr.v30i.2142.

[11] H.-K. Koh, R. Burnasheva, and Y. G. Suh, 'Perceived ESG (Environmental, Social, Governance) and Consumers' Responses: The Mediating Role of Brand Credibility, Brand Image, and Perceived Quality', *Sustainability*, vol. 14, no. 8, p. 4515, Apr. 2022, doi: 10.3390/su14084515.

[12] P. Kumar and S. Sharma, 'Hotel Booking Prediction using Machine Learning', *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 5, pp. 4058–4060, May 2022, doi: 10.22214/ijraset.2022.43036.

[13] E. William Solang, F. Xander Adu, A. Dharma, and N. Gunantara, 'Machine Learning Evaluation for Hotel Cancellation Prediction with Threshold Adjustment and Cost-Based Evaluation', *Indonesian Journal of Machine Learning and Computer Science Institut Riset dan Publikasi Indonesia (IRPI)*, Jan. 17, 2026. doi: https://doi.org/10.57152/malcom.v6i1.2466.

[14] S. E. Whang, Y. Roh, H. Song, and J.-G. Lee, 'Data Collection and Quality Challenges in Deep Learning: A Data-Centric AI Perspective', Dec. 26, 2022, *arXiv*: arXiv:2112.06409. doi: 10.48550/arXiv.2112.06409.

[15] D. N. Ekbote, 'TECHNIQUES OF EXPLORATORY DATA ANALYSIS', vol. 28, no. 2, 2023.