A Comparative Study of Hotel Booking Cancellations through Support Vector Machines and Multilayer Perceptron

Mohammad Talib Akhtar
Mohammad.Akhtar@city.ac.uk

Abstract:
This study looks at how well Support Vector Machines (SVM) and Multilayer Perceptrons (MLP) predict the cancelation of hotel reservations. Both models underwent hyperparameter tweaking and were tuned using an extensive dataset. MLP outperforms SVM in cancelation prediction, according to the performance comparison based on Confusion Matrices and ROC curves. This is a crucial realization for hotel management as it provides a data-driven strategy to reduce cancellation rates and improve overall productivity. Our results highlight the potential of machine learning to improve hospitality services, which represents a major advancement in the industry's use of predictive analytics. In addition to helping to optimize revenue management techniques, this research addresses a significant difficulty in hotel operations and enhances overall customer happiness.

1. Introduction:

Booking cancellations are a persistent concern for the hotel sector, since they may have a big impact on revenue and operational planning. There might be many reasons for canceling hotel booking such as lack of facilities, unattractive offers, unexpected occurrences, obligations, illnesses, accidents, etc [1]. The Support Vector Machine (SVM) and Multilayer Perceptron (MLP), two recent developments in data analytics and machine learning, are used in this work to forecast hotel cancellations. The dataset rich in hotel booking information is used to investigate several model setups and data preparation strategies to increase accuracy. The goal is to use grid search and hyperparameter optimization to find the model that produces the best accurate predictions, and then validate that model using stratified cross-validation.
Some of the previous published works on booking cancellations prediction approach it as a classification problem while most works consider it as a regression problem [2]. We are going to tackle it as a classification problem.
The format of this document is as follows: The dataset is presented in Section 2, together with a description of its features and the preprocessing measures used. The approach, including the model setups and the hyperparameter tuning procedure, is covered in Section 3. A thorough analysis of the findings is provided in Section 4, where the prediction accuracy and other pertinent metrics of the MLP and SVM models are compared for performance. Section 5 brings the study to a close by summarizing the results and outlining potential future research paths for using machine learning to address operational difficulties in the hotel sector.

1.1 Multi Layer Perceptron:

Three different sorts of layers make up a multi-layer perceptron (MLP): input, output, and hidden. The input signal for processing is received by the input layer. The output layer does necessary tasks including categorization and forecasting. The real computing power of the MLP is found in an arbitrary number of hidden layers positioned between the input and output layers. It uses a feature ranking criterion to measure the importance of a feature by computing the aggregate difference, over the feature space, of the probabilistic outputs of the MLP with and without the feature [3].
The MLP's neurons are trained using backpropagation learning. MLPs are designed to approximate any continuous function and can be used to issues that are not linearly separable.

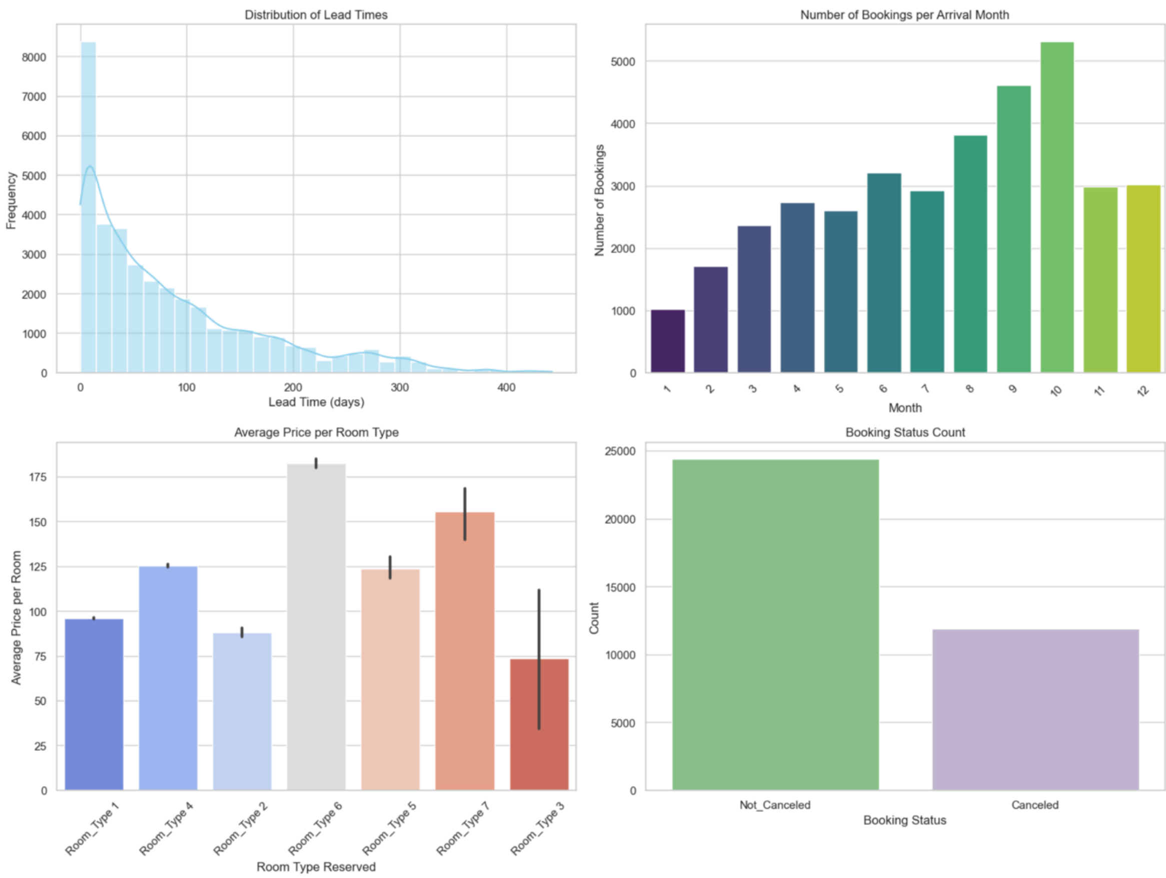
1.2 Support Vector Machine:

Support vector machine, or SVM is the supervised learning algorithm that can be used for both regression and classification problems. However, it is mostly applied to classification problems in machine learning.
The SVM method searches for the optimal line or boundary that could divide n-dimensional space into classes to make it easier to classify new data points in the future
SVM tries to minimize the empirical error while controlling the complexity of the mapping function. This enables it to achieve much better generalization performance on new data [4].

2. Dataset:

The dataset used in this investigation is a subset of 36,275 hotel booking records, comprising 18 independent attributes and a target variable that represents the state of the booking. These features include a plethora of information relevant to hotel reservations, such as guest data, length of stay, booking requirements, and other relevant facts. Predictive modelling that aims to comprehend the causes driving booking cancellations can be based on the binary target variable, "booking status," which distinguishes between cancelled and non-cancelled bookings.
A first examination of the dataset showed a heterogeneous distribution of characteristics, some of which are categorical and so require encoding to numerical forms in order to analyse.

2.1 Initial data analysis:



In the initial analysis of the hotel reservations dataset, several visualizations were generated to explore the underlying patterns and distributions of various features. Figure 1 presents the distribution of lead times, which is notably right skewed, indicating a common preference for making bookings closer to the arrival date. This skewness highlights the trend of last-minute bookings, potentially informing strategies for promotional pricing or availability alerts.

Figure 2 illustrates the monthly booking trends, where a clear seasonality in reservations is observed. The bar chart indicates peaks in certain months, likely reflecting holiday seasons or specific events that drive the higher volume of bookings, which could be pivotal for targeted marketing and resource allocation.

In Figure 3, the average price per room type is depicted, revealing significant variations that could reflect the diversity in room quality, size, or additional amenities offered. The stratified pricing across different room types underscores the value of a differentiated pricing strategy that matches customer expectations and booking patterns.

Lastly, Figure 4 compares the booking statuses, showing a higher prevalence of non-cancelled bookings as opposed to cancelled ones. This contrast suggests a favorable trend in booking retention but also indicates room for improvement in understanding and mitigating the factors leading to cancellations.

3. Methodology:

The methodology for this project was structured to rigorously evaluate and compare the performance of Multilayer Perceptron (MLP) and Support Vector Machine (SVM) models in predicting hotel booking cancellations. To ensure a robust analysis, the original dataset was partitioned into a training set, constituting 80% of the data, and a testing set, making up the remaining 20%.

For the model selection phase, a systematic approach was employed, involving a Grid Search Cross-Validation (GridSearchCV) technique to fine-tune the hyperparameters of both the MLP and SVM models. To achieve a balanced and thorough validation, a stratified k-fold cross-validation strategy was implemented. This technique ensured that each fold had the same proportion of each class label as the full dataset, thus maintaining class distribution and yielding more reliable and unbiased evaluation metrics.

To reduce class imbalance and improve the models' capacity to generalize across different data subsets and avoid overfitting, stratified k-fold cross-validation was essential. This method gave each data point an equal chance of being included in both the training and validation sets. In addition, a grid search technique carefully examined several combinations of hyperparameters to identify the most efficient ones based on validation accuracy. Models that were optimized with these parameters were then retrained using the combined training and validation dataset, which improved their predicted accuracy and made full use of the variety of data that was available.

3.1 Architecture of MLP:

We used a multi-layer perceptron (MLP) neural network as the deep learning algorithm. We experimented with different architectures and hyper-parameters, including the number of hidden layers, the number of neurons in each layer, and the activation function [5].Our Multilayer Perceptron (MLP) architecture is constructed as a pipeline, beginning with a preprocessing step and culminating in a classifier defined by an MLPClassifier from scikit-learn. The MLP features a tiered structure with three hidden layers consisting of 128, 64, and 32 neurons, respectively. This hierarchical arrangement is designed to capture increasingly abstract representations of the input data.

The use of the Rectified Linear Unit (ReLU) activation function in all hidden layers of a model to add non-linearity and address the vanishing gradient issue, aiding in learning complex patterns. Stochastic Gradient Descent (SGD) with backpropagation is used for weight optimization, with an initial learning rate of 0.0001 to ensure gradual convergence and prevent overshooting. The model is trained for 300 epochs to balance learning and avoid overfitting, with a fixed random state for reproducible results. This setup is aimed at enhancing feature learning and making reliable predictions on hotel booking cancellations.

3.2 Architecture of SVM:

The SVM model features a pipeline beginning with data standardization, followed by an RBF kernel-based classifier with a high regularization parameter (C=100) to effectively handle complex hotel reservation data. This setup aims to balance capturing data intricacies and avoiding overfitting, ensuring the model generalizes well to new data.

The SVM model uses a 'gamma' parameter set to 'scale', automatically adjusting to the dataset's complexity and influencing the model's sensitivity to training examples. Along with a fixed random state for consistency, the model's stability is enhanced by its data-driven approach and the RBF kernel's adaptability, making it effective for predicting hotel booking cancellations.

4. Results, Findings & Evaluation:

4.1 Model Selection:

In the model selection phase of our project, we focused on optimizing two sophisticated machine learning models, Multilayer Perceptron (MLP) and Support Vector Machine (SVM), to predict hotel booking cancellations with the highest accuracy possible. Our approach involved extensive hyperparameter tuning using GridSearchCV, a method that systematically explores a range of configurations to identify the most effective model settings.

The MLP model's performance improved from an initial accuracy of 83% to 87% after extensive hyperparameter tuning. Different configurations, including activation functions (ReLU and Tanh), neuron counts in three hidden layers, optimizers (Adam and SGD), L2 regularization strengths, and learning rates, were explored. The optimal setup featured ReLU activation, an L2 regularization of 0.0001, a 128-64-32 neuron structure across the hidden layers, a learning rate of 0.0001, and the Adam optimizer, significantly enhancing the model's predictive accuracy.

Similarly, the SVM model's performance sat at a baseline accuracy of 80%. Through a meticulous parameter search focusing on the regularization parameter C (`'classifier__C': [1, 10, 100]`), the kernel coefficient gamma (`'classifier__gamma': ['scale', 'auto']`), and the choice of kernel (`'classifier__kernel': ['linear', 'rbf']`), we endeavored to enhance the model's predictive power. This thorough tuning process resulted in an optimal configuration of a C value set at 100, gamma at 'scale', and the utilization of the RBF kernel, alluding to a model with a strong preference for regularization that adeptly captures the complex, non-linear intricacies of the hotel booking dataset. The fruits of this laborious tuning were evidenced by an improved accuracy of 87%.

4.2 Algorithm Comparison:

In comparing the performance of our best Multilayer Perceptron (MLP) and Support Vector Machine (SVM) models, we turn to the insights provided by the confusion matrices and ROC curves. Examining the confusion matrices (Fig 1 for MLP and Fig 2 for SVM), the MLP model exhibits a propensity for a higher number of false negatives, incorrectly predicting 456 positive cases as negative, as opposed to SVM's 395. Conversely, the SVM model has a slightly lower count of false positives at 575 compared to the MLP's 792, suggesting a conservative nature in predicting the positive class.

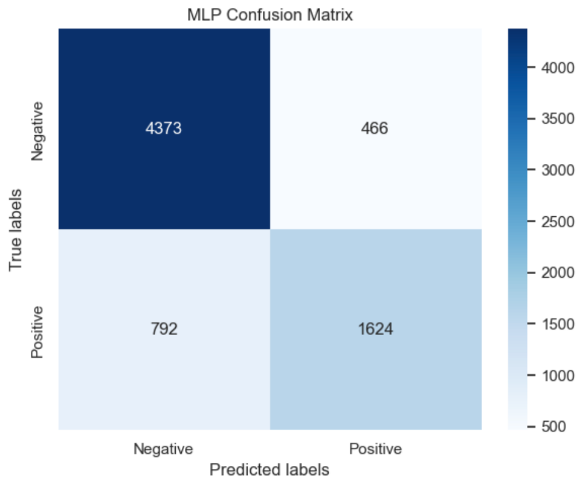


Figure 1

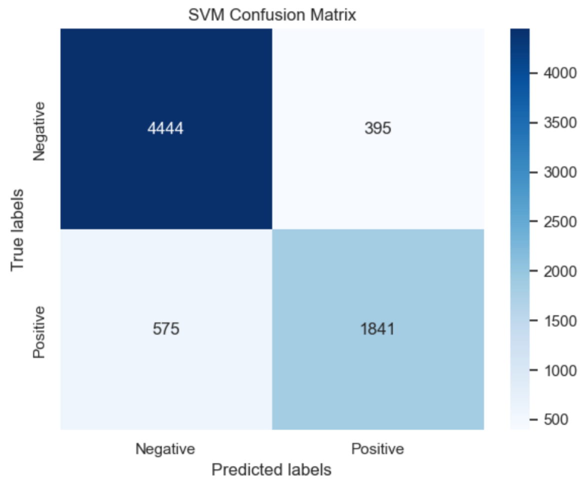


Figure 2

Based on the provided ROC (Receiver Operation Curve) curves for both MLP and SVM models, there's a notable difference in their performance on the training and test datasets. For the training data, the MLP model shows an AUC of 0.97, which is slightly higher than the SVM's AUC of 0.87, indicating a marginally better fit to the training dataset.

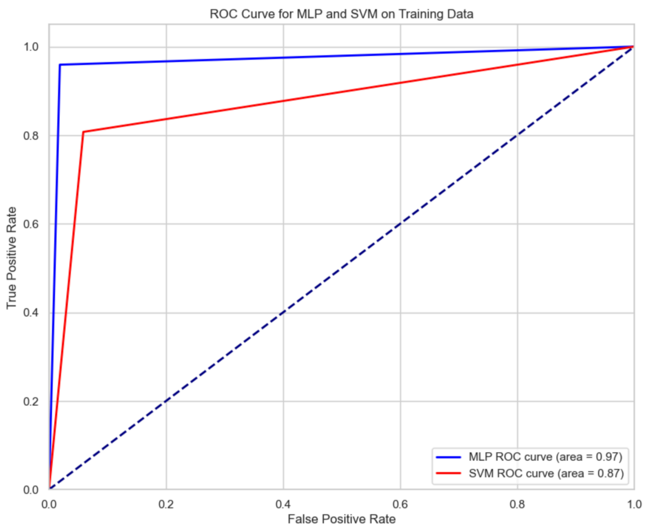


Figure 3

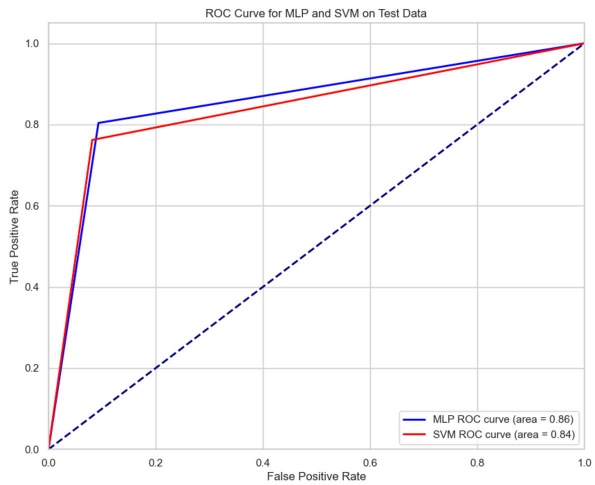


Figure 4

On the test data, the MLP model maintains a lead with an AUC of 0.86 compared to the SVM's AUC of 0.84, suggesting that the MLP model generalizes slightly better to unseen data. However, both models display strong predictive abilities and very similar performance characteristics, with only marginal differences between their ROC curves. The higher AUC values for both models on the training data compared to the test data suggest potential overfitting, which is a common occurrence in machine learning models.

When it comes to model selection, our MLP achieved an impressive 87% accuracy with optimal parameters including 'relu' activation, a three-layer architecture, and 'adam' solver. In contrast, the SVM model, fine-tuned with a 'rbf' kernel, 'C' value of 100, and 'gamma' set to 'scale', proved to be exceptionally well-suited for the task at hand. Though the exact accuracy for SVM wasn't specified, the configuration suggests a model that can handle the intricacies of the dataset effectively.

In summary, the choice between MLP and SVM would depend on the specific costs associated with false positives and false negatives within the context of hotel booking cancellations. The ROC curve analysis aids in visualizing the trade-offs between sensitivity and specificity, guiding stakeholders in selecting the model that aligns best with their operational objectives and risk preferences.

5. Conclusion:

Throughout this project, several key lessons have been learned that are invaluable for future machine learning endeavors. Firstly, the importance of hyperparameter tuning in the performance of models like MLP and SVM cannot be overstated; it greatly affects their predictive capabilities. We also learned that the initialization of neuron weights in MLP can lead to variability, which underscores the need for careful consideration of random states for reproducibility in results. MLP has emerged as a transformative power in the hotel industry, offering the potential to revolutionize various fields of its operations and customer experiences [5].
For future work, exploring a broader set of machine learning models and techniques, such as ensemble methods, could potentially enhance predictive accuracy and model robustness. This could include bagging or boosting methods that might reduce the sensitivity to specific training dataset partitions and neuron weight initializations.
Additionally, we see potential in investigating feature importance and extraction to understand better which attributes most significantly impact booking cancellations. By diving deeper into the dataset, we can uncover more nuanced relationships within the features and perhaps streamline the models to focus on the most predictive indicators.

6. References

1. M. S. Satu, K. Ahammed and M. Z. Abedin, "Performance Analysis of Machine Learning Techniques to Predict Hotel booking Cancellations in Hospitality Industry," 2020 23rd International Conference on Computer and Information Technology (ICCIT), DHAKA, Bangladesh, 2020, pp. 1-6, doi: 10.1109/ICCIT51783.2020.9392648
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9392648&isnumber=9392586

2. N. Antonio, A. de Almeida and L. Nunes, "Predicting Hotel Bookings Cancellation with a Machine Learning Classification Model," 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 2017, pp. 1049-1054, doi: 10.1109/ICMLA.2017.00-11.
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8260781&isnumber=8260597

3. J. -B. Yang, K. -Q. Shen, C. -J. Ong and X. -P. Li, "Feature Selection for MLP Neural Network: The Use of Random Permutation of Probabilistic Outputs," in IEEE Transactions on Neural Networks, vol. 20, no. 12, pp. 1911-1922, Dec. 2009, doi: 10.1109/TNN.2009.2032543.
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5282531&isnumber=5338643

4. X. Yuan, Z. Yang, G. Zouridakis and N. Mullani, "SVM-based Texture Classification and Application to Early Melanoma Detection," 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, USA, 2006, pp. 4775-4778, doi: 10.1109/IEMBS.2006.260056.
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4462869&isnumber=4461641

5. A. R. Yadav, S. Shekhar, A. Vidyarthi, R. Prakash and R. Gowri, "Hyper-Parameter Tuning with Grid and Randomized Search Techniques for Predictive Models of Hotel Booking," 2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM), Roorkee, India, 2023, pp. 1-5, doi: 10.1109/ELEXCOM58812.2023.10370718.
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10370718&isnumber=10370033

Intermediate Results:

When evaluating the performance of machine learning models like Support Vector Machines (SVM) and Multi-Layer Perceptrons (MLP), an accuracy metric of 80% for SVM and 83% for MLP indicates how often the models correctly predicted the outcomes compared to the actual labels in the test set.

SVM employs a linear kernel with a regularization parameter C that is set to 1.0. Selecting a linear kernel implies that the model is intended to identify a linear hyperplane that divides the classes. an SVM with a linear kernel can withstand overfitting, particularly in high-dimensional fields; the regularization parameter C aids in balancing the need to retain a low generalization error while obtaining a low training error.

On the other hand, MLPs architecture contains three hidden layers (128, 64, and 32 neurons each) and the ReLU activation function, in addition to the SGD solver. The MLP can more successfully identify non-linear correlations in the data with this arrangement. The model has been set to converge smoothly across a comparatively high number of epochs. This might result in a stronger generalization on unobserved data. The reason for the MLP's better accuracy of 83% might be attributed to its depth and capacity to understand intricate patterns.

Glossary:

Feedforward neural network: An artificial neural network in which there is no cycle in the connections between the nodes is called a feedforward neural network. Recurrent neural networks, on the other hand, have connections that can loop back on themselves.

Backpropagation: Backpropagation in a Multi-Layer Perceptron (MLP) is a training algorithm used to minimize the error by adjusting the weights of the network. It works by propagating the error backward through the network, from the output layer to the input layers, updating weights along the way based on the gradient of the error.

Robust Analysis: Robust analysis refers to the evaluation of a system or model's ability to perform consistently under varying conditions or when faced with uncertainties, anomalies, or deviations in input data.

ReLU (Rectified Linear UnitReLU is an activation function that produces zero output in the absence of a positive input and outputs the input when it is positive in neural networks.

SGD (Stochastic Gradient Descent): SGD is an optimization method used in training neural networks. At each iteration SGD modifies the model's parameters (weights) using a subset (batch) of the data, making the learning process faster and more memory-efficient.

Learning Rate: The step size for each iteration of the optimization process is determined by the learning rate (like SGD), affecting how quickly a model converges to the minimum loss.

Gradual Convergence: Gradual convergence refers to the iterative process of optimization in machine learning, where the model incrementally improves its accuracy over iterations until it reaches or approaches an optimal state of minimal error.

C (Regularization Parameter): The regularization parameter "C" in support vector machines (SVM) regulates the trade-off between lowering model complexity for improved generalization to new data and attaining a low error on the training set.

Kernel: In SVM, the kernel function converts the input data into a higher-dimensional space so that classes may be more successfully divided using a hyperplane. Radial basis function (RBF), polynomial, and linear kernels are examples of common kernels.

Gamma (γ): Used specifically with the RBF kernel in SVMs, gamma indicates how far a single training sample may impact the model; low values indicate "far," while high values indicate "close."

Confusion Matrix: It’s a table that displays a classification model's performance on a set of test data with known real values. True positives, false positives, true negatives, and false negatives are all counted.

ROC (Receiver Operating Characteristic) Curve: Plotting the true positive rate (sensitivity) against the false positive rate (1-specificity), the ROC curve shows how diagnostically capable a binary classifier system is when its discrimination threshold is changed.

AUC (Area Under the Curve): The area beneath the ROC curve is measured by AUC, or area under the curve. AUC gives an overall performance metric over all potential classification thresholds; a larger AUC denotes better model performance.