Hotel Reservation Cancellations: Predicting Customer Churn with Explainable AI

Introduction

A hotel, fully booked for peak season, faces last-minute cancellations from guests who booked months in advance. This creates chaos with vacant rooms, lost revenue, and scrambling staff. The challenge is to predict which bookings are likely to cancel in advance to prevent this.

Problem Definition:

- **Who** are the guests most likely to cancel?
- **What** booking behaviours increase the likelihood of a cancellation?
- **When** are these cancellations most likely to occur?
- **Where** do most cancellations come from?
- **Why** are guests cancelling?

Proposed Solution:

We propose a data-driven approach that uses machine learning and Explainable AI (XAI) to accurately predict hotel reservation cancellations. Our solution not only forecasts which bookings are most likely to cancel but also explains why, helping hotels take proactive measures, reduce financial losses, optimize resource allocation, and enhance guest satisfaction.

Project Motivation

Our project is motivated by the growing unpredictability and financial strain in the hospitality industry caused by hotel chains facing high sunken costs from operating above necessary capacity. As the industry grows, mitigating costs from sudden cancellations is critical.

Literature Review:

Research by **Sánchez et al. (2020)** highlights that critical impact that last minute cancellations tend to have on hotel revenues. This study leverages Artificial Intelligence techniques to predict which customers are most likely to cancel close to their stay with an accuracy of over 80%.

Similarly, **Lin (2023)** uses a combination of random forests and logistic regression to forecaster cancellations. These models focus on optimising hotel operations, but like many machine learning approaches, they function as "black boxes" providing little to no insight into the decision making processes behind their predictions.

Özkurt (2024) addresses this gap by introducing LIME and SHAP, two prominent XAI techniques, to interpret the predictions of machine learning models. By integrating explainable AI into our project, this model will not only contribute significantly to the comprehension of customer churn in the hospitality industry but will also offer insights crucial for prevention strategies for hotels.

Additionally, research from **Antonio (2017)** underscores the importance of hotels predicting cancellations to increase revenue, making substantial impact on their business.

Dataset Description

Dataset Link

https://www.kaggle.com/datasets/saadharoon27/hotel-booking-dataset

Dataset Description

We chose the Hotel Booking Dataset from Kaggle for its comprehensive features on hotel reservations, including guest demographics, booking details, and cancellation statuses. Key features like lead time, previous cancellations, deposit type, and booking channels provide insights into booking behavior, helping assess cancellation risk. By focusing on these indicators, we aim to identify patterns contributing to cancellations.

Intended Results

Our goal is to develop a model that predicts customer churn, helping hotels manage cancellations, optimize revenue, reduce operational costs, and improve guest satisfaction. We focus on reducing resource waste from overbooking or under-utilization while ensuring transparency, fairness, and avoiding bias in our model. Previous research has predicted hotel churn with success, and we aim for at least 85% accuracy, using Explainable AI to enhance insights. Key metrics for evaluation include:

Success Metrics:

- Classification report including precision, recall, f1-score
- Train and Test Accuracy
- Train and Test area under the ROC curve (AUC)

eXplainable Artificial Intelligence (XAI) metrics:

- LIME (Local Interpretable Model-agnostic Explanations)
- SHAP (SHapley Additive exPlanations)

Methods

Data Preprocessing

Prior to model training, we undertook the following steps to prepare the raw hotel booking dataset. This step involved multiple rounds of systematic data cleaning, preprocessing, and feature engineering. First, the dataset was thoroughly analysed to evaluate which columns were categorical, numerical, and which columns had the maximum number of missing values/needed cleaning. After that, the data preprocessing was broken down into the following steps:

- **Data Privacy:** removed sensitive personal information including names, emails, phone numbers, credit card details.
- **Data Transformation:** we converted the hotel type into binary format, representing ‘City Hotel’ as 0 and ‘Resort Hotel’ as 1.
- **Feature Engineering:** Extracted a day of the week (Monday = 0; Sunday = 6). Created a binary feature indicating weekend arrivals.
- **Missing and invalid values’ data handling:** filled in missing values in the ‘children’ column with the median value. Removed rows with an adult count of zero, as these are considered invalid.
- **Room Type Change Indicator:** Created a feature to indicate if room type changed between booking and check-in. Dropped original room type columns after creating the indicators.
- **Encoding for Categorical Variables:** one-hot encoding for standard categorical variables. Frequency encoding for ‘country’, ‘agent’, ‘company’.
- **Outlier Management:** limited extreme values in ‘adr’ (average daily rate) column to manage outliers effectively.
- **Reservation Timing:** computed number of days between last reservation status date and arrival date (temporarily computing by combining arrival day number, month number, and year number). Dropped original reservation status data after calculation. XGBoost does not handle datetime values, and hence the columns split into day number, month number, and year number were kept that way and not combined into a datetime format.
- **Numerical Standardisation:** standardised to mean=0 and standard deviation=1. This ensures consistency and compatibility across features. After the steps above were implemented, the processed dataset was split into training and testing sets (80:20 split) for model training and evaluation.

Important Additional Preprocessing Step:

Initially, the model achieved an unusually high accuracy of 100%. To investigate this, we examined the correlation between each feature and the target variable (is_cancelled). This revealed that 3 features (status_Canceled, status_Check-Out, and days_since_reservation_status) had a strong correlation with the target. These features were inadvertently providing information about the cancellation status directly, causing data leakage and artificially inflating the model's performance. Therefore, these 3 columns were dropped.

Machine Learning Algorithm/Models:

Algorithm 1: XGBoost

XGBoost, which stands for Extreme Gradient Boosting, is a scalable, distributed gradient-boosted decision tree (GBDT) machine learning algorithm. It enhances predictive accuracy by iteratively improving upon decision trees within a gradient-boosting framework. We chose XGBoost as the initial model in our project to predict hotel churn, as it offers flexibility, computational efficiency, and strong performance with classification tasks. Its ability to handle various data types, coupled with its robust design to limit overfitting, made it an ideal choice for identifying bookings most likely to cancel.

Algorithm 2: Multilayer Perceptron (MLP)

A Multilayer Perceptron is an artificial neural network that employs non-linear activation functions to interpret and learn complex patterns in data, meaning that it can understand non-linear relationships as well. This makes it a useful machine learning model for our task.

Algorithm 3: LightGBM

Light Gradient Boosting Machine is a state-of-the-art gradient boosting framework based on tree-based machine learning models that is suitable for classification tasks, thus making it an optimal choice for our project.

Algorithm 4: Logistic Regression

We chose logistic regression as one of the models for this project as it is a classification method that uses a sigmoid function that takes the input as independent variables and outputs discrete probability values of 0 and 1, which fits the classification nature of this task.

Algorithm 5: CatBoost

CatBoost is a machine learning algorithm that relies on gradient boosting to handle categorical data, such as the data that is concerning this project. It's designed specifically for quantitatively classified tasks that has the ability to not only handle categories but even numerical features. It also has an interesting in-built function that other algortihms discussed so far do not - the ability to handle different missing values in a dataset in a manner that can reduce overfitting and improve how the dataset performs when put through training and testing periods.

Results and Discussion

Algorithm 1: XGBoost

1a) Classification report, Accuracy, AUC

	Dataset	Accuracy	AUC
0	Train	0.898055	0.966399
1	Test	0.884318	0.955350

```
Test Set Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.92      0.91     14951
           1       0.86      0.82      0.84      8847

    accuracy                           0.88     23798
   macro avg       0.88      0.87      0.88     23798
weighted avg       0.88      0.88      0.88     23798
```

A high training accuracy of 89.81% and AUC of 0.966 indicates XGBoost’s effective learning on the training data. The test accuracy of 88.43% and an AUC of 0.955 suggests that the model generalises well to unseen data with minimal overfitting. This slight drop in test accuracy and AUC compared to training metrics is expected and is within a reasonable range. All in all, These results imply that the model captures underlying patterns effectively while maintaining robustness when applied to new data.

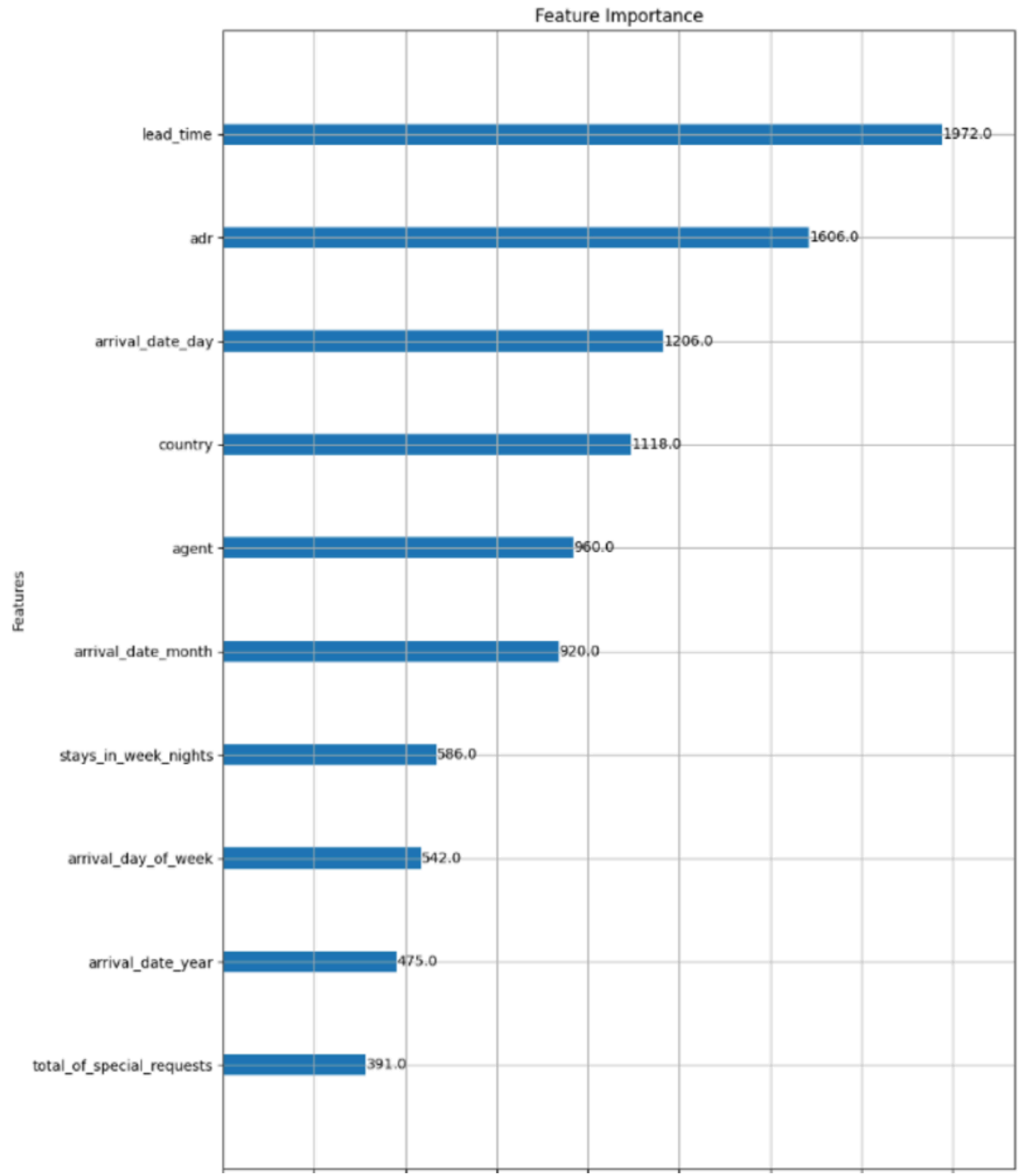
A closer look at AUC: AUC (Area Under the Curve) is a metric used to evaluate the performance of binary classification models. AUC refers to the area under the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (Sensitivity) against the False Positive Rate (1 - Specificity) at various threshold settings. An AUC score closer to 1 demonstrates the model's ability to correctly rank positive samples higher than negative samples. In our case, a high test AUC of 0.955 shows that the model is effective at distinguishing between cancellation and non-cancellations.

	Dataset	Accuracy	AUC
0	Train	0.898055	0.966399
1	Test	0.884318	0.955350
2	Train (Optimized)	0.901407	0.969262
3	Test (Optimized)	0.886713	0.957812

```
Test Set Classification Report (Optimized Model):
              precision    recall  f1-score   support

           0       0.90      0.92      0.91     14951
           1       0.87      0.82      0.84      8847

    accuracy                           0.89     23798
   macro avg       0.88      0.87      0.88     23798
weighted avg       0.89      0.89      0.89     23798
```

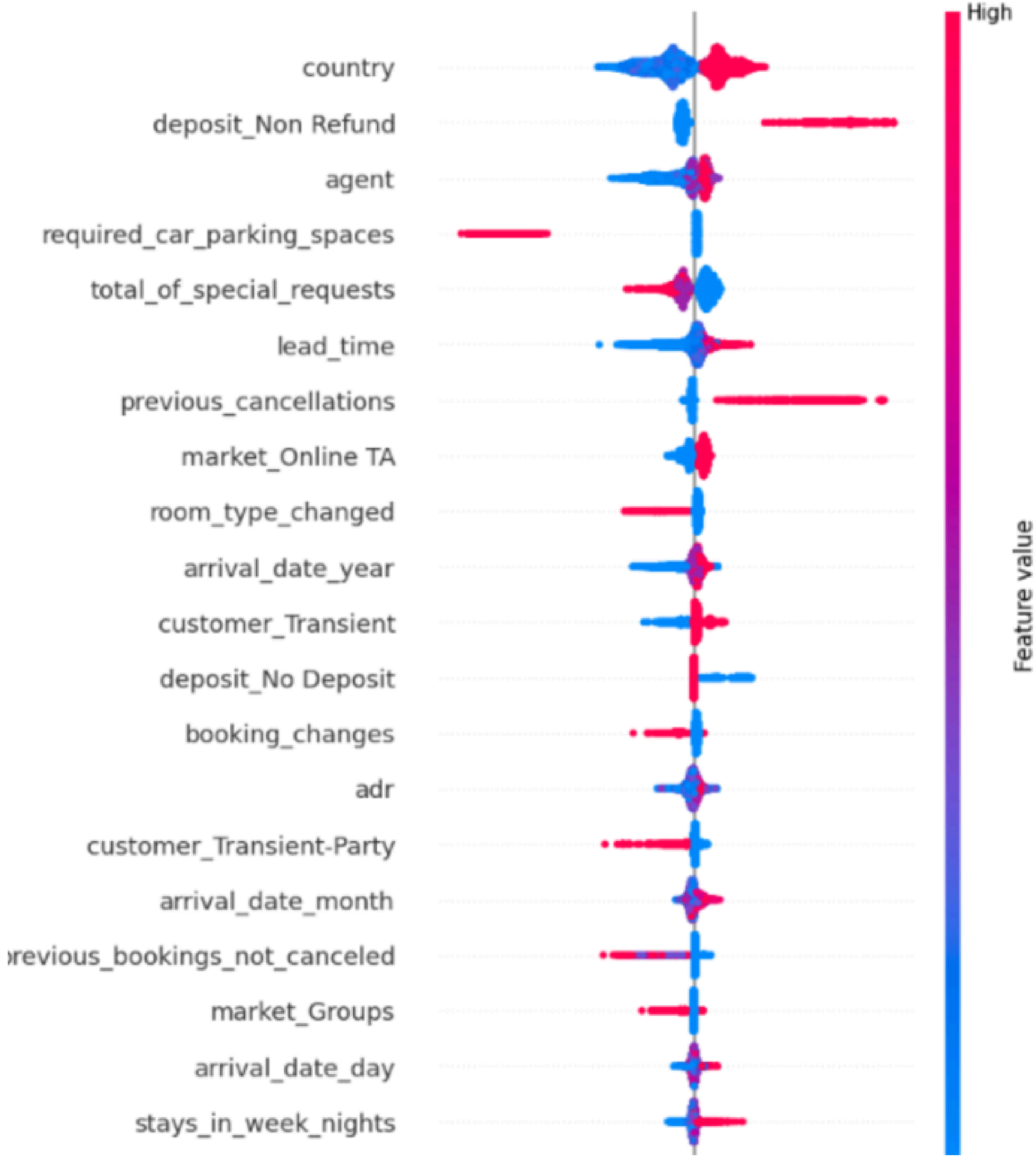
Post hyperparameter tuning, there was a negligible change in accuracy, which increased for the training set from 0.88 to 0.90 and the test set from 0.88 to 0.89. This reinforces that the initial configuration of the parameters was already close to optimal for the baseline model. The AUC remained stable for both the training and test remained stable at 0.966 and 0.95, reinforcing that the baseline model was already distinguishing between the classes effectively.

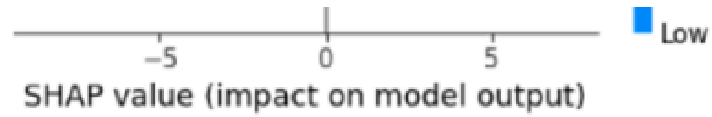
1b) Feature Importance


Feature Importance
Features
lead_time
adr
arrival_date_day
country
agent
arrival_date_month
stays_in_week_nights
arrival_day_of_week
arrival_date_year
total_of_special_requests
1972.0
1606.0
1206.0
1118.0
960.0
920.0
586.0
542.0
475.0
391.0

The feature importance plot illustrates the 10 most influential features that helped in the XGBoost model's predictions Lead Time had the highest importance with a score of 1972. This indicates that the number of days between the booking date and the check-in date significantly affects the prediction as longer lead times might be correlated with cancellations since guests have more opportunities to change their plans. The Average Daily Rate ranked second in importance with a score of 1606 as higher ADR values might cause guests to cancel their booking due to financial reasons. Arrival Date Day is the third most important feature with a score of 1206 as specific days in a month could correlate with peak travel dates, special events, or general booking tendencies, influencing cancellation probabilities. Some other important features in descending order are country, agent, and arrival_data_month which have feature importance scores of 1118, 960 and 920 respectively.

1c) SHAP


High
country
deposit_Non Refund
agent
required_car_parking_spaces
total_of_special_requests
lead_time
previous_cancellations
market_Online TA
room_type_changed
arrival_date_year
customer_Transient
deposit_No Deposit
booking_changes
adr
customer_Transient-Party
arrival_date_month
revious_bookings_not_canceled
market_Groups
arrival_date_day
stays_in_week_nights
Feature value

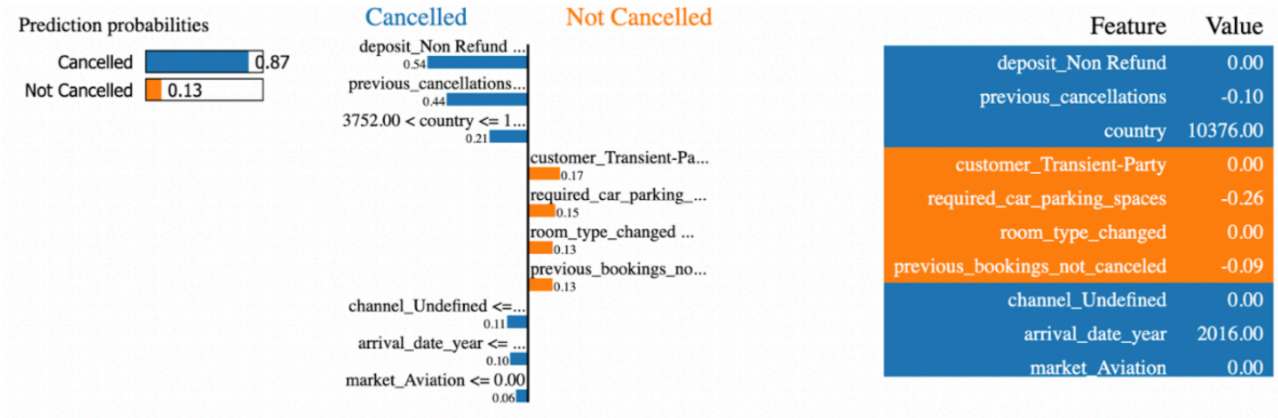


This SHAP summary plot illustrates the various impacts of features on the likelihood of calculation in this predictive model. The features are ordered by their overall impact on the prediction, with the most influential at the top. Based on our previous inference regarding positive and more negative SHAP values, we can see that high features (shown on the plot in red) and low features (shown on the plot in blue) show different influences across the various SHAP values. For example, high values of “deposit_non_refund” and “previous_cancellations” seem to increase the risk of cancellation as shown by the high feature red on the positive side of the graph. On the other hand, features such as "total_of_special_requests" and "required_car_parking_spaces" tend to decrease the likelihood of cancellation, as indicated by the concentration of high feature values (red) on the negative side of the graph. Customers with non-refundable deposits or a history of cancellations are more prone to canceling, while those with specific requests or parking requirements appear more committed to their bookings.

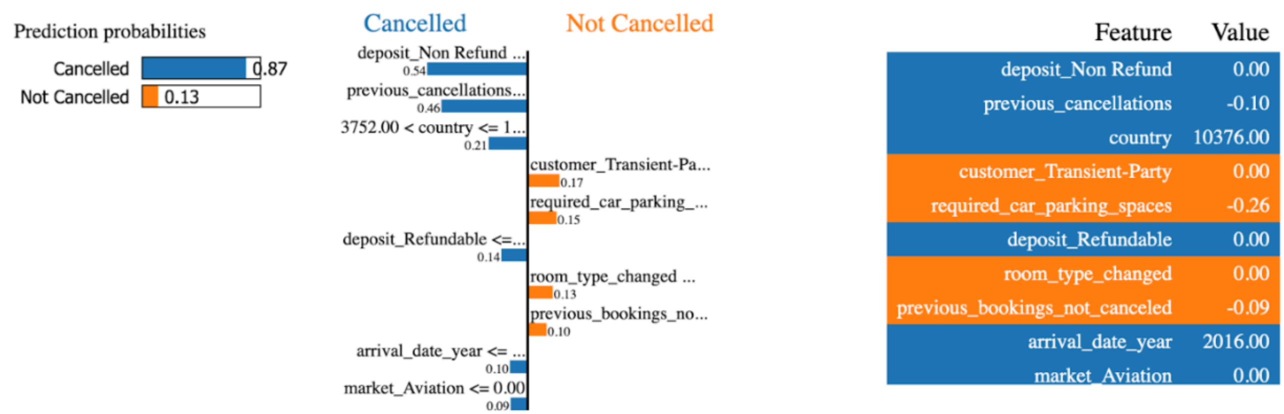
One key observation is the significant impact of the "deposit_Non Refund" feature, which shows a strong positive correlation with cancellations. This could be attributed to the fact that a non-refundable deposit often indicates a smaller upfront payment. Customers who make such small deposits may feel less financially committed to their booking and are more likely to cancel when plans change. Similarly, "previous_cancellations" also shows a strong positive influence, indicating that customers with a history of cancellations are more prone to repeating this behavior.

Conversely, features like "total_of_special_requests" and "required_car_parking_spaces" exhibit a negative relationship with cancellations. These indicate a higher level of planning and commitment on the part of the customer, as specific requests or parking needs suggest a stronger intention to follow through with the booking. The presence of these factors likely reduces the likelihood of cancellations, showcasing how customer behavior patterns influence predictive outcomes.

1d) LIME



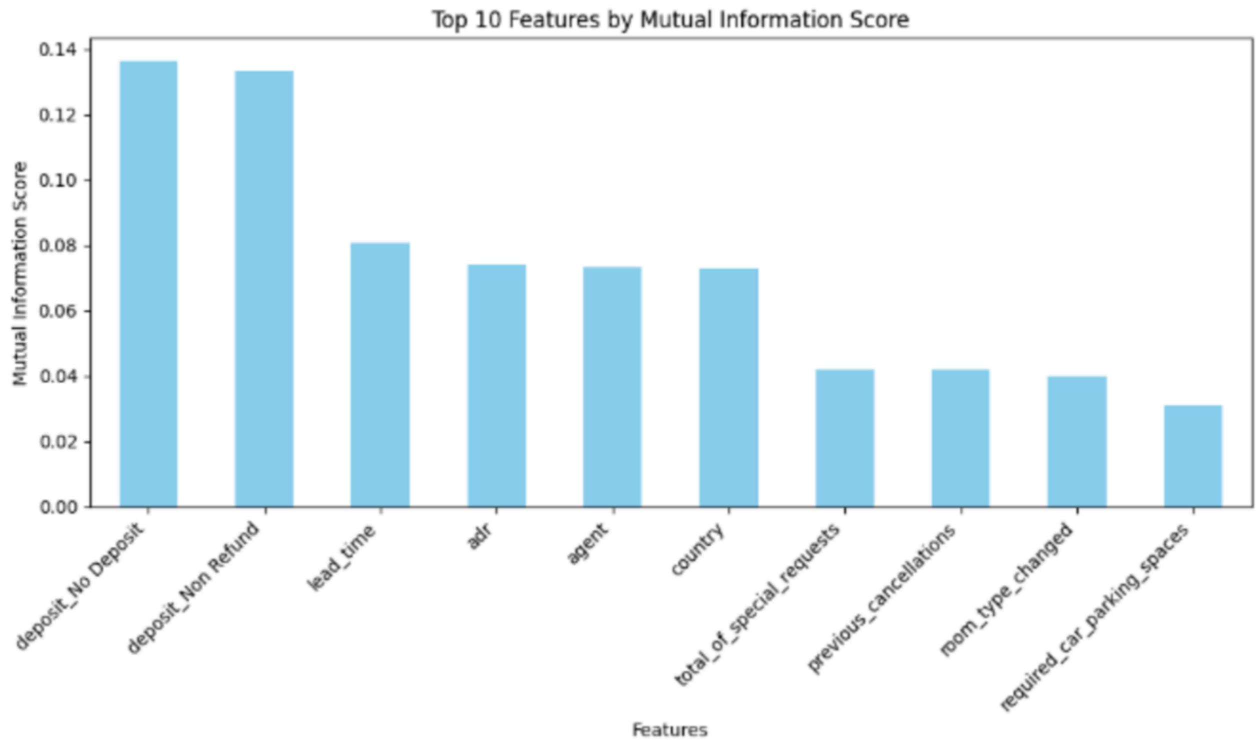
The LIME visualization for the training data provides an insight into the impact of features on the likelihood of a booking being classified as "Cancelled" or "Not Cancelled." The most influential feature for predicting cancellations was deposit_Non_Refund, which significantly increases the likelihood of cancellation as signified by the high positive SHAP value of 0.54. Similarly, a higher count of previous_cancellations strongly correlates with an increased probability of cancellation, as expected. Additionally, countries with certain values are also associated with a higher cancellation likelihood, reflecting possible geographic trends and proximity to the location. On the other hand, features like required_car_parking_spaces and previous_bookings_not_canceled contribute to reducing the likelihood of cancellation. These features suggest that bookings with prior successful completions or certain logistical requirements are less prone to cancellations. Lesser influences, such as room_type_changed, customer_Transient-Party, and channel_Undefined, show minimal contributions to the classification outcome.



According to the visualization for the testing data as shown above, the most influential feature for predicting cancellations was deposit_Non_Refund with a LIME value of 0.54, which significantly increases the likelihood of cancellation. Previous_cancellations trails behind closely with a positive LIME value of 0.46, signifying a strong correlation with an increased probability of cancellation, as expected. Additionally, countries with certain values are also associated with a higher cancellation likelihood, reflecting possible geographic trends or proximity to the location of stay.

On the other hand, features like customer_Transient_-Party and required_car_parking_spaces contribute to reducing the likelihood of cancellation with LIME values of 0.17 and 0.15 respectively. Milder influences, such as room_type_changed and previous_bookings_not_canceled have LIME scores of 0.13 and 0.10, emerging as important factors that indicate cancellations.

1e) Mutual Information



This chart highlights the Top 10 Features most strongly related to the prediction outcome based on their Mutual Information (MI) scores. The higher the MI score, the more informative the feature is in predicting cancellations.

Deposit_No Deposit and Deposit_Non Refund are, in order, the two features that have the highest mutual information scores, both close to 0.14, indicating their significant influence on the target variable. Whether a deposit is required or not, and the type of deposit (non-refundable), strongly correlates with booking behavior and cancellation likelihood. Lead_time ranked third in the plot with an MI score of about 0.08 highlighting that the time gap between booking and check-in played a critical role in predicting cancellations. A longer lead time might indicate a higher chance of cancellation due to greater uncertainty over time. The ADR (Average Daily Rate) emerged in fourth place with a score slightly below that of lead_time, reflecting the financial concerns of the guests. Higher rates might lead to cancellations in case the guests reconsider their expenditures. Some other features that have comparable MI scores are adr, agent, and country with MI scores close to 0.08.

Algorithm 2: Multilayer Perceptron (MLP)

2a) Classification report, Accuracy, AUC

```
Test Set Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.90      0.87     14951
           1       0.82      0.71      0.76      8847

    accuracy                           0.83     23798
   macro avg       0.83      0.81      0.82     23798
weighted avg       0.83      0.83      0.83     23798
```

Initial Model Performance:

	Dataset	Accuracy	AUC
0	Train	0.832165	0.909756
1	Test	0.832381	0.910423

Prior to hyperparameter tuning, the model achieved an accuracy of 83.22% on the test dataset and 83.22% on the training dataset. The AUC score was slightly above 0.91 for both the train and test datasets. These metrics indicate that the initial model performed reasonably well in classifying the dataset, though there was slight room for improvement, particularly in terms of recall for the minority class (label 1).

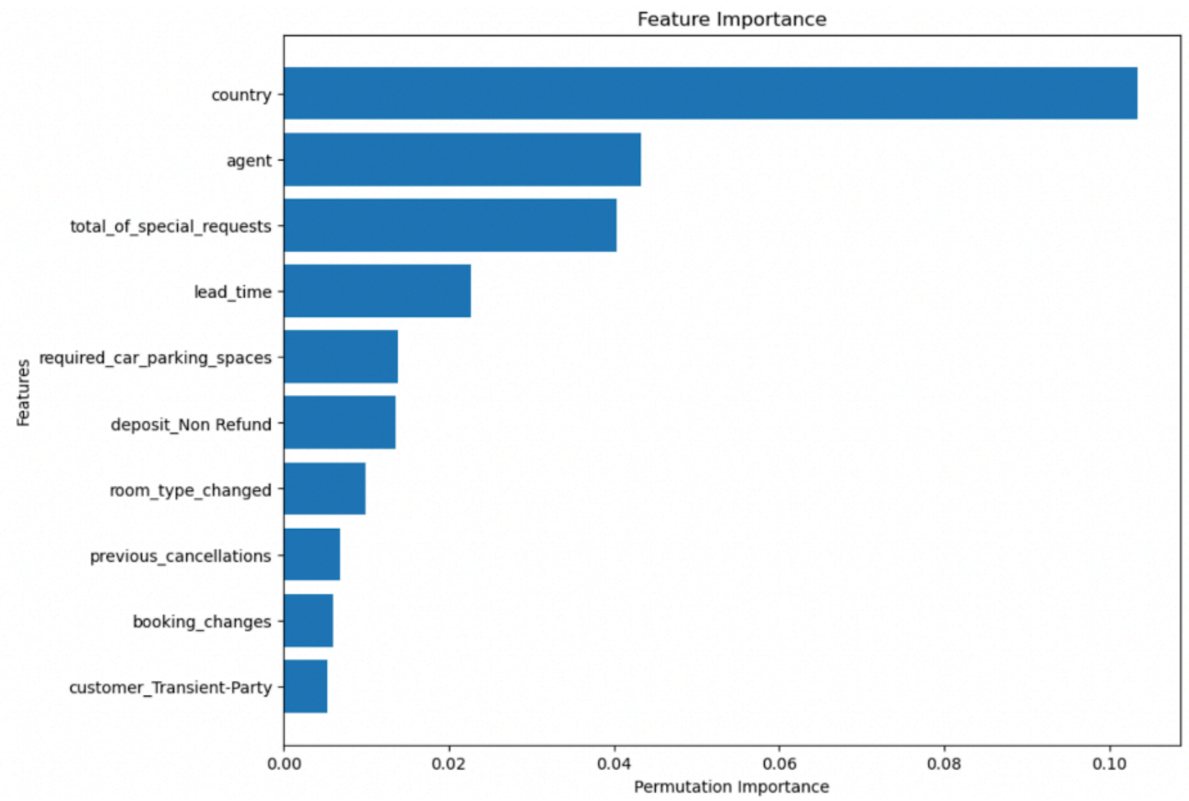
```
Test Set Classification Report (Optimized Model):
              precision    recall  f1-score   support

           0       0.84      0.90      0.87     14951
           1       0.82      0.71      0.76      8847

    accuracy                           0.83     23798
   macro avg       0.83      0.81      0.82     23798
weighted avg       0.83      0.83      0.83     23798
```

	Dataset	Accuracy	AUC
0	Train	0.832165	0.909756
1	Test	0.832381	0.910423
2	Train (Optimized)	0.832165	0.909756
3	Test (Optimized)	0.832381	0.910423

After optimization, the model only demonstrated modest gains in the classification and AUC metrics. The training and test accuracies remained stable at 83.22%, and the AUC scores also remained consistent at approximately 0.91. The consistency in metrics between the pre- and post-optimization models demonstrates that the initial model was robust and generalizable.

2b) Feature Importance

Feature importance, also known as permutation feature importance in MLP models, is a method used to figure out how important each feature is for a model's predictions. It works by taking one feature at a time and randomly shuffling its values, breaking its connection to the target variable. Then, the model's performance is checked to see how much it drops. If the model's performance gets worse, it means that the feature is important for making accurate predictions.

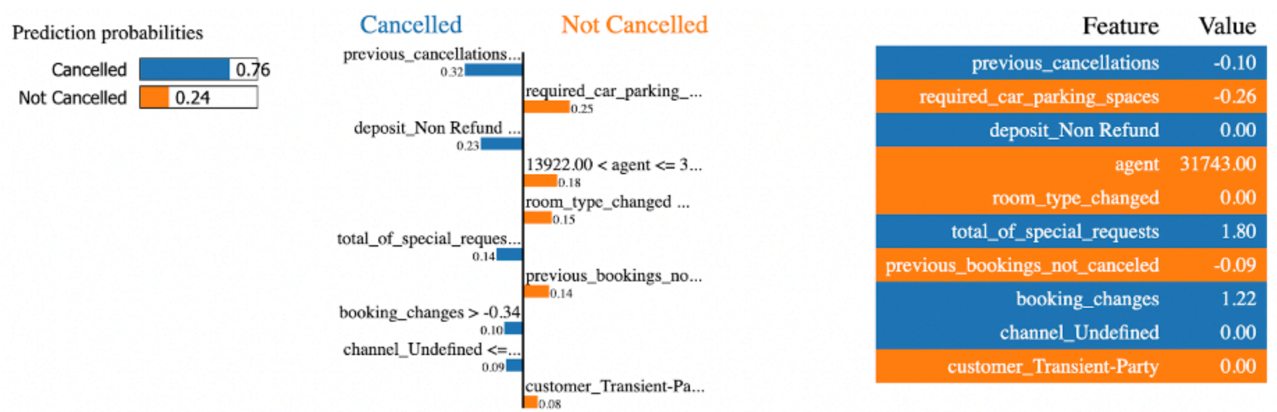


As visualized in the plot above, 'country' is the most important feature, demonstrating the highest permutation importance score. This indicates that geographical and regional variations significantly affect the model's predictive ability due to their impact on cancellation. The next most impactful features are 'agent' and 'total_of_special_requests', which are also of high importance. 'Agent' reflects the role of intermediate booking platforms and middlemen whereas 'total_of_special_requests' captures how customer preferences and complexities in their bookings, both of which are shown to be strong predictors of cancellations. 'Lead_time' is also a significant predictor, suggesting that the duration between booking and the actual stay significantly influences cancellation behavior.

2c) SHAP

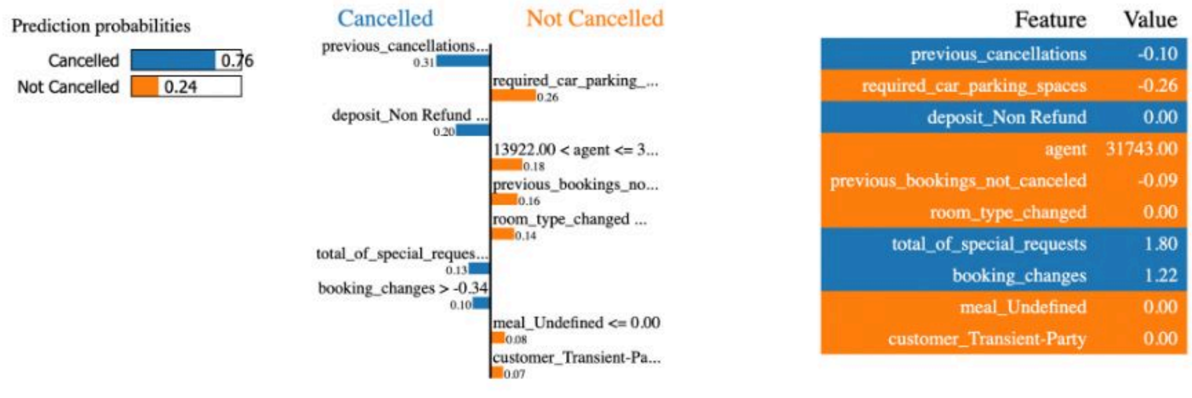
According to the SHAP plot above, **"country"** was the most important feature for the MLP model, demonstrating a wide range of SHAP values. This means that the impact of **"country"** on the model's predictions varies significantly depending on the specific country. High values (red points) tend to skew predictions positively, suggesting some countries may have a higher likelihood of cancellations, while others (blue points) contribute less consistently to cancellations. **"Deposit_Non Refund"** is another feature that has a positive influence on cancellation probabilities when its value is high. This suggests that non-refundable deposits strongly increase the likelihood of cancellations, likely due to the higher financial risk associated with such bookings. On the contrary, low values have less impact on cancellation likelihood. The feature **"agent"** shows a moderate spread, with high values (red) leaning towards a positive contribution to cancellations.**"Lead_time"**, which represents the time between booking and check-in, has a rather predictable trend where higher values (red points) contribute positively to cancellation likelihood as it can be inferred that customers who book far in advance might be more likely to change their plans and cancel. On the other hand, shorter lead times (blue points) have a neutral or slightly negative impact on cancellation, reflecting more definitive booking behavior.

2d) LIME



For the first instance of the training dataset, "previous_cancellations" is the most influential feature driving cancellations, with a significant contribution of 0.32. This highlights that customers with a history of cancellations remain highly likely to cancel again. "deposit_Non_Refund" also plays a substantial role (0.23), indicating that non-refundable deposits correlate with higher cancellation probabilities, likely due to customer hesitance. "total_of_special_requests" (0.14) suggests that customers making numerous specific requests are more prone to cancellations, potentially due to unmet expectations. Meanwhile, "booking_changes > -0.34" and "channel_Undefined" have smaller but notable effects, where frequent booking changes and undefined channels increase the likelihood of cancellation.

In contrast, the most important features for non-cancellations include "required_car_parking_spaces" (0.25), which suggests stronger commitment when parking is requested, and "agent", indicating that bookings through certain agents are less likely to be canceled. "total_of_special_requests" (1.80) also appears in the non-cancellation context, reflecting its nuanced role in both increasing and decreasing cancellation risks depending on specific circumstances.

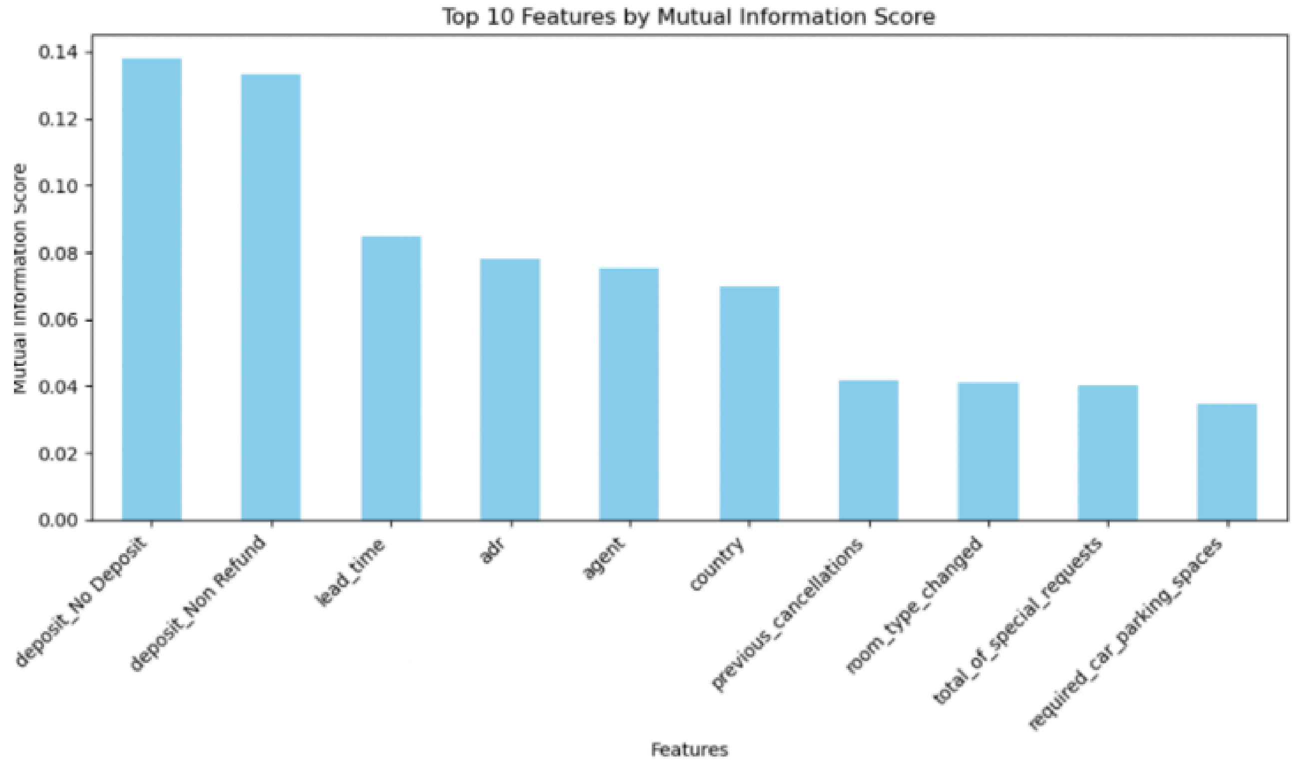


In general, trends in the training dataset were reflected in the LIME values for the testing dataset. Mirroring its role in the training dataset, "previous_cancellations" was the dominant feature (0.31) influencing cancellations in the testing dataset as well. This demonstrates that the model generalizes the significance of cancellation history across unseen data well. "deposit_Non_Refund" (0.20) and "total_of_special_requests" (0.13) continue to show similar patterns, reinforcing that they are strong predictors.

However, features such as "booking_changes > -0.34" and "channel_Undefined" contribute slightly less in the testing data, suggesting slight variability in their role.

For non-cancellations, "required_car_parking_spaces" was also consistent as the top feature (0.26), again reinforcing its stabilizing role. "agent" (value >13,922) and "room_type_changed" (0.14) all maintain their importance, demonstrating consistency in how agents and flexible bookings reduce cancellation risks.

2e) Mutual Information



According to the plot above, deposit type features such as "deposit_No_Deposit" and "deposit_Non_Refund" have the highest MI scores, suggesting that the type of deposit has a significant influence on the target outcome. "No Deposit" scenarios might indicate lower customer commitment, whereas "Non-Refundable" deposits may increase the likelihood of cancellations due to cost-sensitive decision-making. Lead Time ranks third, emphasizing the importance of how far in advance a booking is made. Longer lead times might be correlated with higher cancellation rates, possibly due to increased uncertainty over time. ADR (Average Daily Rate) also has a strong MI score, indicating that the pricing structure plays a key role. Higher rates might be associated with cancellations for budget-conscious customers. The "Agent" feature, which refers to booking intermediaries, appears next. Certain agents might have patterns associated with cancellations, such as targeting a demographic more likely to cancel.

Algorithm 3: LightGBM

3a) Classification report, Accuracy, AUC

In order to evaluate LightGBM's performance, we looked at the accuracy and AUC metrics of the model, both prior and post optimization. The AUC metric provides crucial insights into the model's ability to classify bookings as either "Cancelled" or "Not Cancelled" while also measuring its robustness in discerning between these two classes across different decision thresholds.

```
Test Set Classification Report:
              precision    recall  f1-score   support

           0       0.89      0.92      0.90     14951
           1       0.86      0.80      0.83      8847

    accuracy                           0.88     23798
   macro avg       0.87      0.86      0.87     23798
weighted avg       0.88      0.88      0.88     23798
```

	Dataset	Accuracy	AUC
0	Train	0.882182	0.955921
1	Test	0.877595	0.952725

Before optimization, the LightGBM model demonstrated a strong performance. The training accuracy was recorded at 88.22%, with a testing accuracy of 87.76%. This close alignment between training and testing accuracies indicated that the model was generalizing well to unseen data without much overfitting. The training set also achieved a high AUC score of 0.9559, while the testing set followed closely at 0.9527. The high AUC values suggest that the model effectively distinguished between the canceled and not canceled classes.

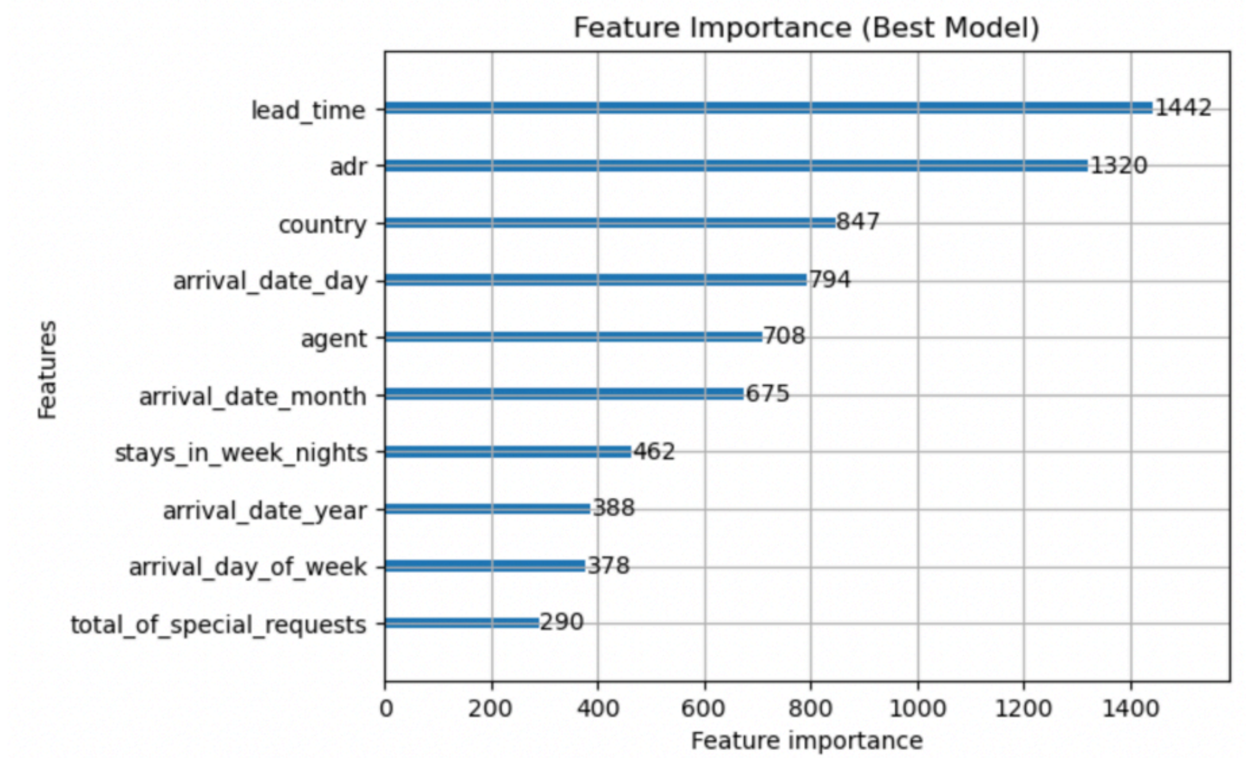
```
Test Set Classification Report (Optimized Model):
              precision    recall  f1-score   support

           0       0.90      0.93      0.91     14951
           1       0.87      0.83      0.85      8847

    accuracy                           0.89     23798
   macro avg       0.88      0.88      0.88     23798
weighted avg       0.89      0.89      0.89     23798
```

	Dataset	Accuracy	AUC
0	Train	0.882182	0.955921
1	Test	0.877595	0.952725
2	Train (Optimized)	0.902919	0.970849
3	Test (Optimized)	0.888898	0.958883

After conducting hyperparameter tuning via a gridsearch search, where different thresholds for parameters such as learning rate, maximum depth, and boosting rounds were explored, the model's performance showed MINOR some improvement. The training accuracy increased to 90.29%, while the testing accuracy rose to 88.89%. Similarly, the AUC metrics demonstrated an increase, with the training set achieving a score of 0.9708 and the testing set improving to 0.9588. While this enhancement reflects a more precise ability to rank positive and negative samples correctly, the improvement is not significant and signifies that the original configuration of the baseline model was already close to optimal.

3b) Feature Importance

For LightGBM, feature importance is based on the trained model where each feature’s importance is calculated based on the gain achieved in the accuracy due to the feature when splitting decision trees.



1.Most Important Features:

- lead_time: This feature has the highest importance, aligning with its consistent impact in SHAP analysis. Customers with longer lead times are more prone to cancellations.
- adr: The second most important feature, reflecting the impact of higher booking costs on customer decisions to cancel.
- country: Geographic differences remain a significant factor in predicting cancellations, as seen in both SHAP and feature importance analyses.

2. Moderately Important Features:

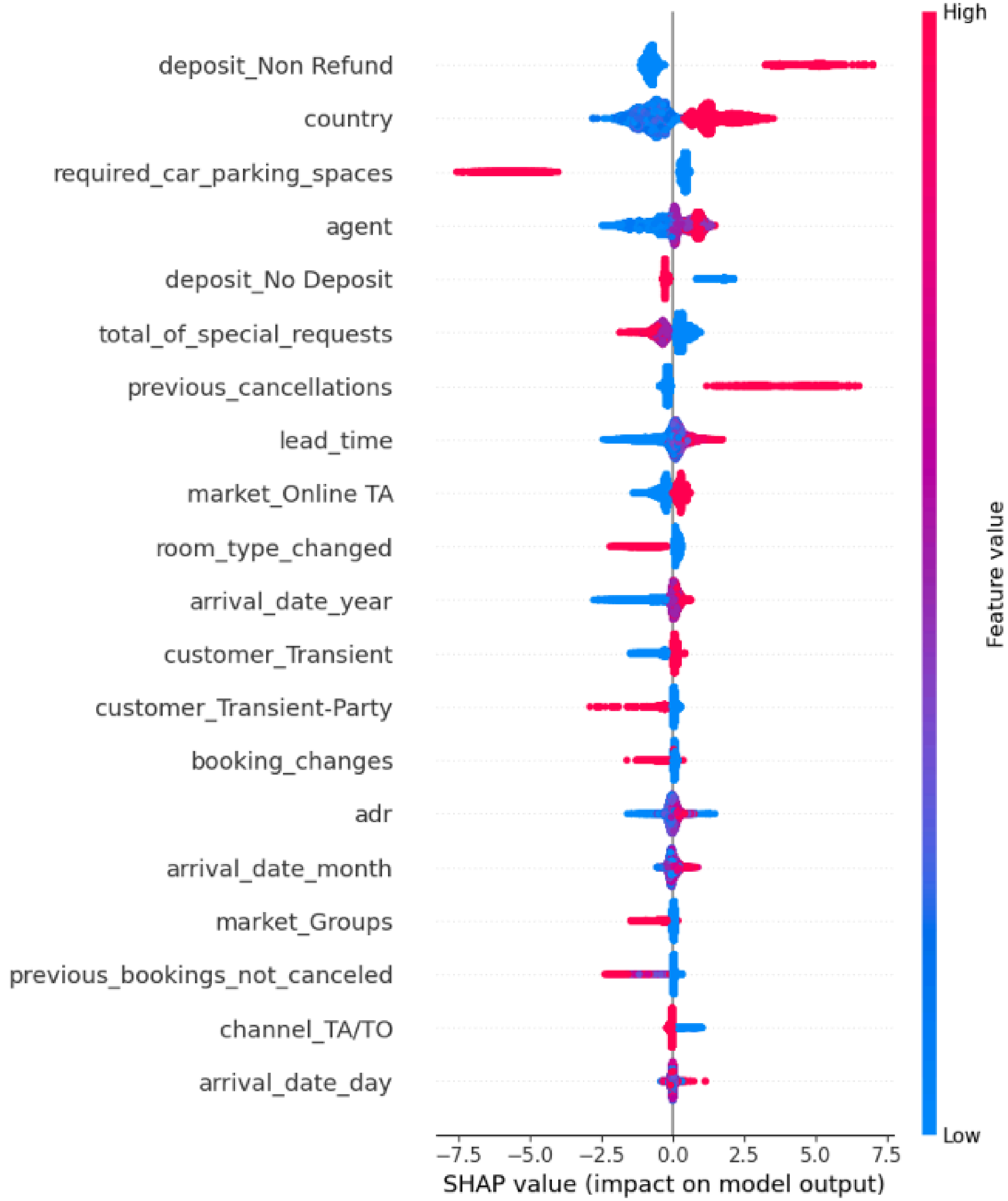
- arrival_date_day and arrival_date_month: These features capture booking timing and its relationship with cancellations.

- agent: Booking agents influence cancellations, likely reflecting differences in policies or customer demographics across platforms.

3. Commitment-Related Features:

- stays_in_week_nights: Longer stays during the week correlate with fewer cancellations, signifying the customer’s commitment.
- total_of_special_requests: This feature continues to demonstrate its importance in reducing cancellations, signifying the customer engagement.

3c) SHAP


deposit_Non Refund
country
required_car_parking_spaces
agent
deposit_No Deposit
total_of_special_requests
previous_cancellations
lead_time
market_Online TA
room_type_changed
arrival_date_year
customer_Transient
customer_Transient-Party
booking_changes
adr
arrival_date_month
market_Groups
previous_bookings_not_canceled
channel_TA/TO
arrival_date_day
−7.5 −5.0 −2.5 0.0 2.5 5.0 7.5
SHAP value (impact on model output)
High
Feature value
Low

Feature Impact on Cancellation Likelihood:

- deposit_Non_Refund: This feature is the most impactful, with higher values (shown in red) significantly increasing the likelihood of cancellation. Customers with non-refundable deposits may be more likely to cancel despite the penalties due to unforeseen circumstances or misaligned expectations.
- country: The country of origin shows a broad range of SHAP values, indicating its variable impact across different regions. Some countries are associated with higher cancellation rates, while others exhibit reduced probabilities.
- required_car_parking_spaces: Parking requests (blue) consistently reduce the likelihood of cancellation. This feature reflects a higher commitment to travel plans.

Behavioral and Financial Indicators:

- lead_time: Longer lead times (red) increase the likelihood of cancellations, as customers with more time between booking and arrival are more prone to changing plans. Conversely, shorter lead times (blue) are associated with fewer cancellations.
- adr (Average Daily Rate): Higher booking costs contribute to increased cancellations, suggesting customers may cancel when finding more affordable alternatives.

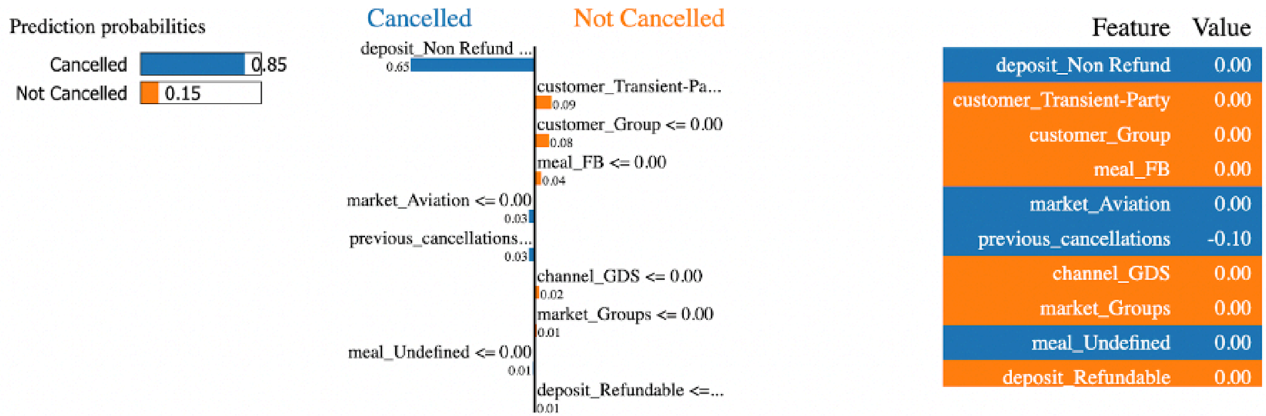
Commitment and Flexibility:

- total_of_special_requests: Customers who make special requests (blue) are less likely to cancel, indicating a higher level of investment in their bookings.
- room_type_changed: Room type changes contribute positively to cancellation likelihood, potentially reflecting dissatisfaction or indecision during the booking process.

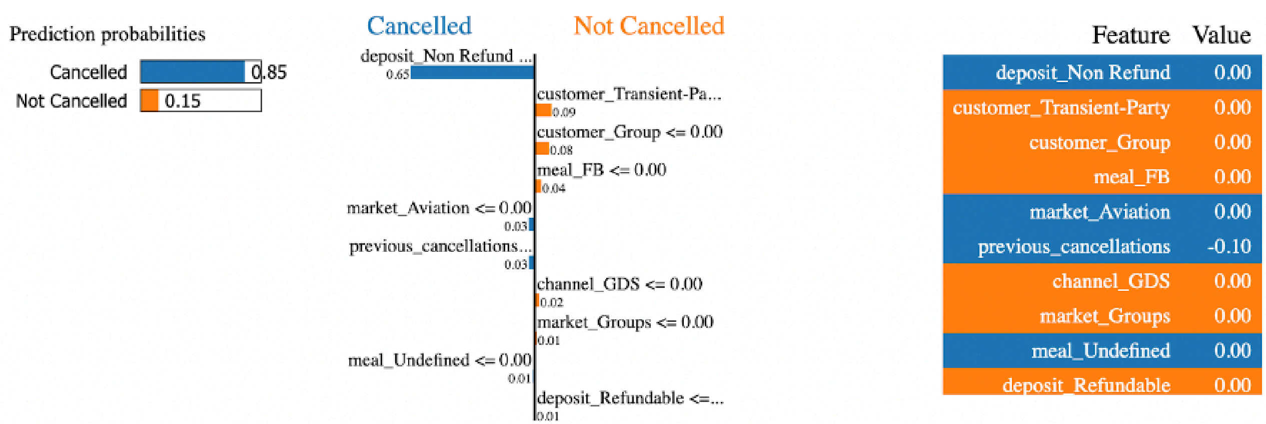
Booking Channels and Agents:

- agent: The booking agent significantly impacts predictions. Specific agents or booking platforms might correlate with particular cancellation behaviors.
- market_Online_TA: Online travel agency bookings contribute to cancellations, reflecting the ease of canceling through such platforms.

3d) LIME

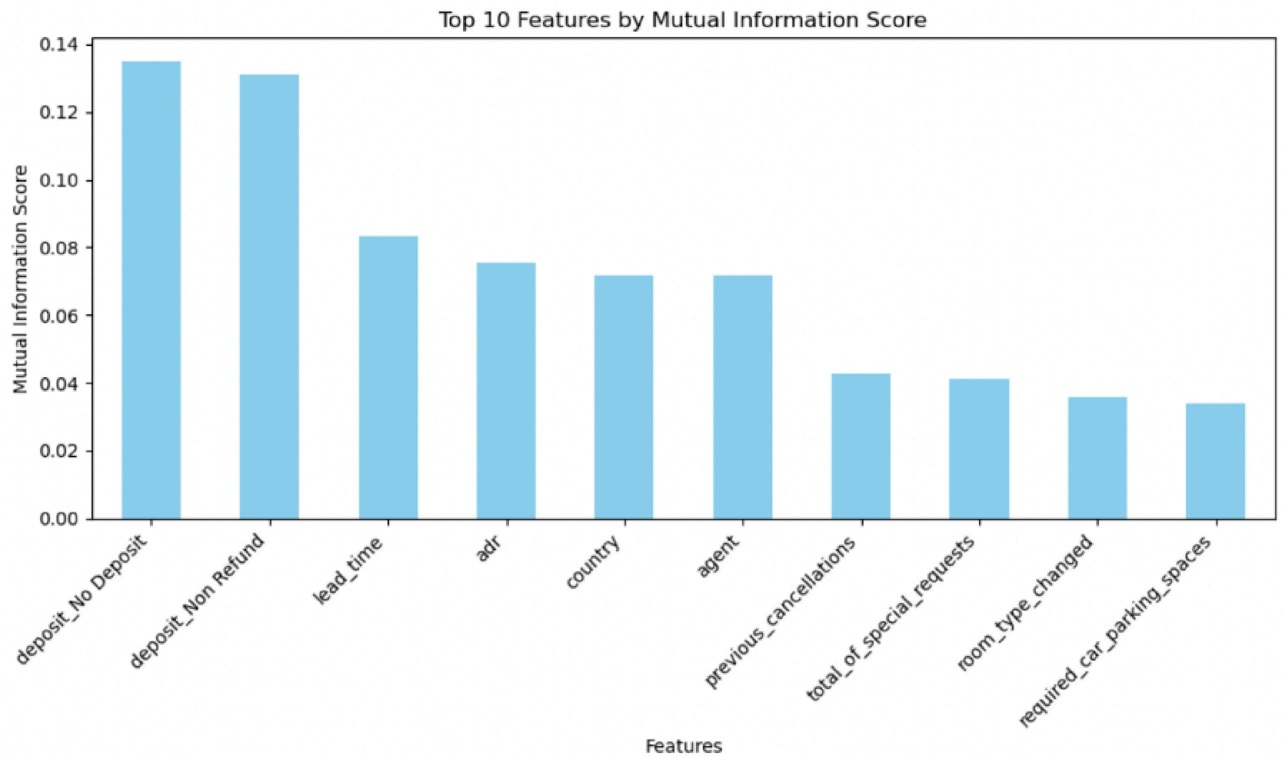


The visualization highlights the optimized LightGBM’s model's prediction that a booking is Cancelled with a high probability of 85%, compared to a 15% likelihood for Not Cancelled for the first instance of the training data. The feature contributions show that the "Cancelled" prediction is strongly influenced by the deposit_Non Refund feature, with a significant positive contribution of 0.65, indicating its strong association with cancellations. Other features like market_Aviation and previous_cancellations also contribute marginally to the "Cancelled" class. Conversely, the "Not Cancelled" prediction is influenced by features such as customer_Transient-Party and customer_Group, though their contributions are relatively smaller. Overall, the prediction is dominated by features indicative of cancellation behavior, emphasizing the impact of non-refundable deposits on the likelihood of cancellation.



The visualization above of the optimized LightGBM model estimates an 85% likelihood of cancellation and 15% chance of not being canceled for the first instance of the testing data. The dominant feature influencing the cancellation class is the deposit_Non Refund feature with a weight of 0.64, emphasizing the strong influence of non-refundable deposits on cancellation decisions. Other minor contributions to cancellation come from features such as previous_cancellations and market_Aviation, although their impact is minimal. On the other hand, the "Not Cancelled" prediction is influenced by features like customer_Transient-Party and deposit_Refundable, which slightly favor the "Not Cancelled" outcome but are outweighed by the dominant cancellation drivers. Overall, the classifier leans toward Cancelled due to the combined effect of these features.

3e) Mutual Information



1. High Mutual Information Features:

- deposit_Non_Refund and deposit_No_Deposit: These features dominate the MI rankings, emphasizing the critical role of deposit types in cancellation behavior.
- lead_time: Longer lead times increase cancellation probabilities, aligning with trends observed in SHAP analysis.

2. Behavioral Trends

- previous_cancellations: A history of cancellations strongly predicts future cancellations, highlighting the importance of customer behavior patterns.
- total_of_special_requests: Special requests reduce cancellation likelihood, suggesting higher customer engagement and commitment.

3. Booking and Demographic Factors

- agent: Booking agents influence cancellations.
- country: The country of origin significantly affects cancellation rates, as seen in both SHAP and MI analyses.

Algorithm 4: Logistic Regression

4a) Classification report, Accuracy, AUC

We evaluated the performance of our baseline logistic regression model using performance and AUC metrics which highlight the mode's robustness in distinguishing between the canceled and not_canceled classes. The following are the results before optimizing the parameters:

Test Set Classification Report:

	precision	recall	f1-score	support
0	0.83	0.91	0.86	14951
1	0.81	0.68	0.74	8847
accuracy			0.82	23798
macro avg	0.82	0.79	0.80	23798
weighted avg	0.82	0.82	0.82	23798

```
Initial Model Performance:
```

	Dataset	Accuracy	AUC
0	Train	0.820504	0.899902
1	Test	0.821666	0.901848

The accuracy metric, representing the percentage of correctly classified samples, is consistent across the training and testing datasets. For the training set, the accuracy was measured at 82.05%, while for the testing set, it was slightly higher at 82.17%. This alignment between training and testing accuracy suggests that the model generalizes well to unseen data without much overfitting.

Furthemore, the training AUC was recorded at 0.8999, and the testing AUC was slightly higher at 0.9018. The high AUC values demonstrate that the unoptimized model can effectively rank positive and negative samples across various thresholds, making it a reliable tool for classification tasks. The similarity in AUC values between training and testing datasets further supports the model's generalizability.

```
Test Set Classification Report (Optimized Model):
              precision    recall  f1-score   support

           0       0.83      0.91      0.86     14951
           1       0.81      0.67      0.74      8847

    accuracy                           0.82     23798
   macro avg       0.82      0.79      0.80     23798
weighted avg       0.82      0.82      0.82     23798
```

```
            Dataset  Accuracy       AUC
0             Train  0.820504  0.899902
1              Test  0.821666  0.901848
2  Train (Optimized)  0.820399  0.899963
3   Test (Optimized)  0.821666  0.901944
```

The optimized version of the Logistic Regression model demonstrated minimal improvement in performance metrics, as the accuracy for the training set was recorded at 82.04%, while the accuracy for the testing set remained consistent at 82.17%, suggesting that the default parameters were already close to the optimal for this dataset.

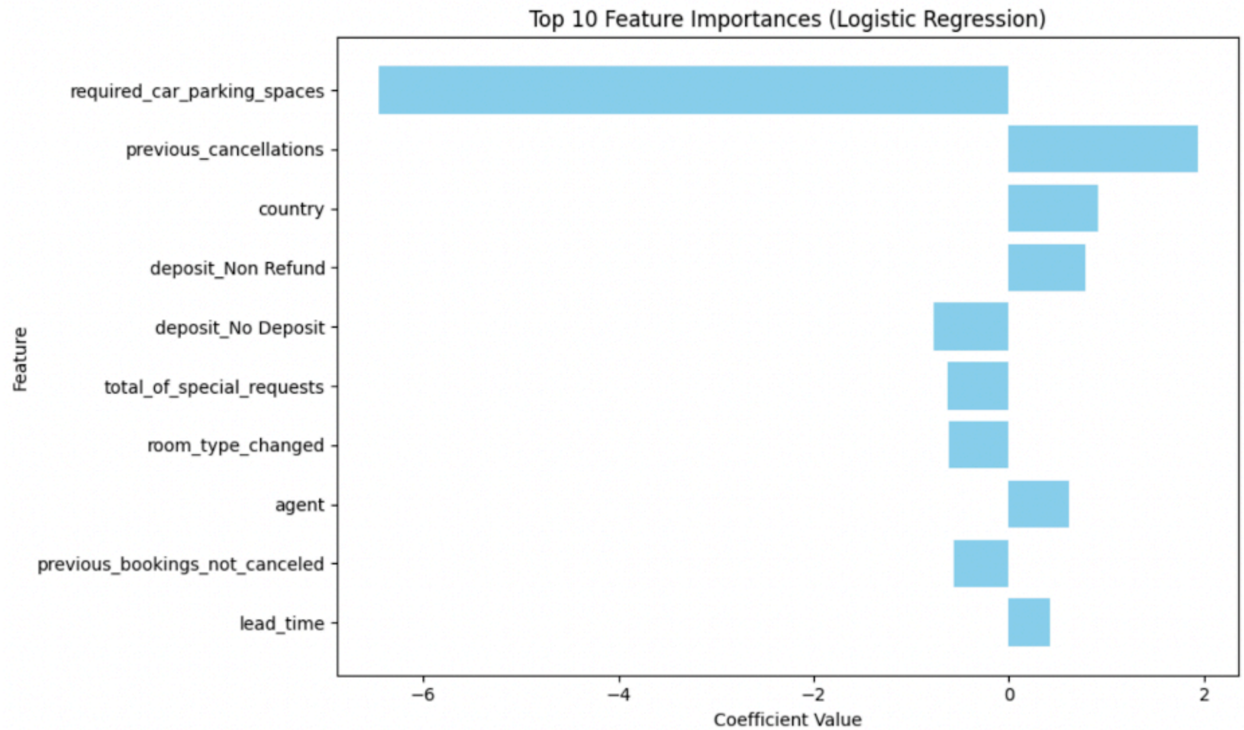
The AUC metrics also remained largely unchanged post-optimization. The training AUC showed a minor improvement from 0.899902 to 0.899963, while the testing AUC saw a slight increase from 0.901848 to 0.901944. These marginal differences suggest that the default Logistic Regression parameters already captured the dataset's underlying patterns effectively.

4b) Feature Importance

In logistic regression, the magnitude of the coefficient directly corresponds with its importance to the model. The greater the magnitude, the higher its importance.

Each coefficient reflects the impact of a one-unit change in the predictor variable on the log odds of the outcome, assuming all other variables remain constant.

- Positive Coefficient: A positive value indicates that as the predictor variable increases, the likelihood of the positive outcome also increases.
- Negative Coefficient: A negative value suggests that as the predictor variable increases, the likelihood of the positive outcome decreases.



1.Most Influential Features:

- required_car_parking_spaces: This feature has the highest negative coefficient, indicating it strongly reduces the likelihood of cancellation. Parking requests are a reliable indicator of customer commitment to their travel plans.
- previous_cancellations: This feature has a strong positive coefficient, reflecting its direct correlation with increased cancellations.
- deposit_Non_Refund: This feature contributes positively to cancellations, indicating customers with non-refundable deposits are more likely to cancel, despite the financial penalty.
- deposit_No_Deposit: Conversely, this feature negatively correlates with cancellations, showing that customers without upfront financial commitments are less likely to cancel.

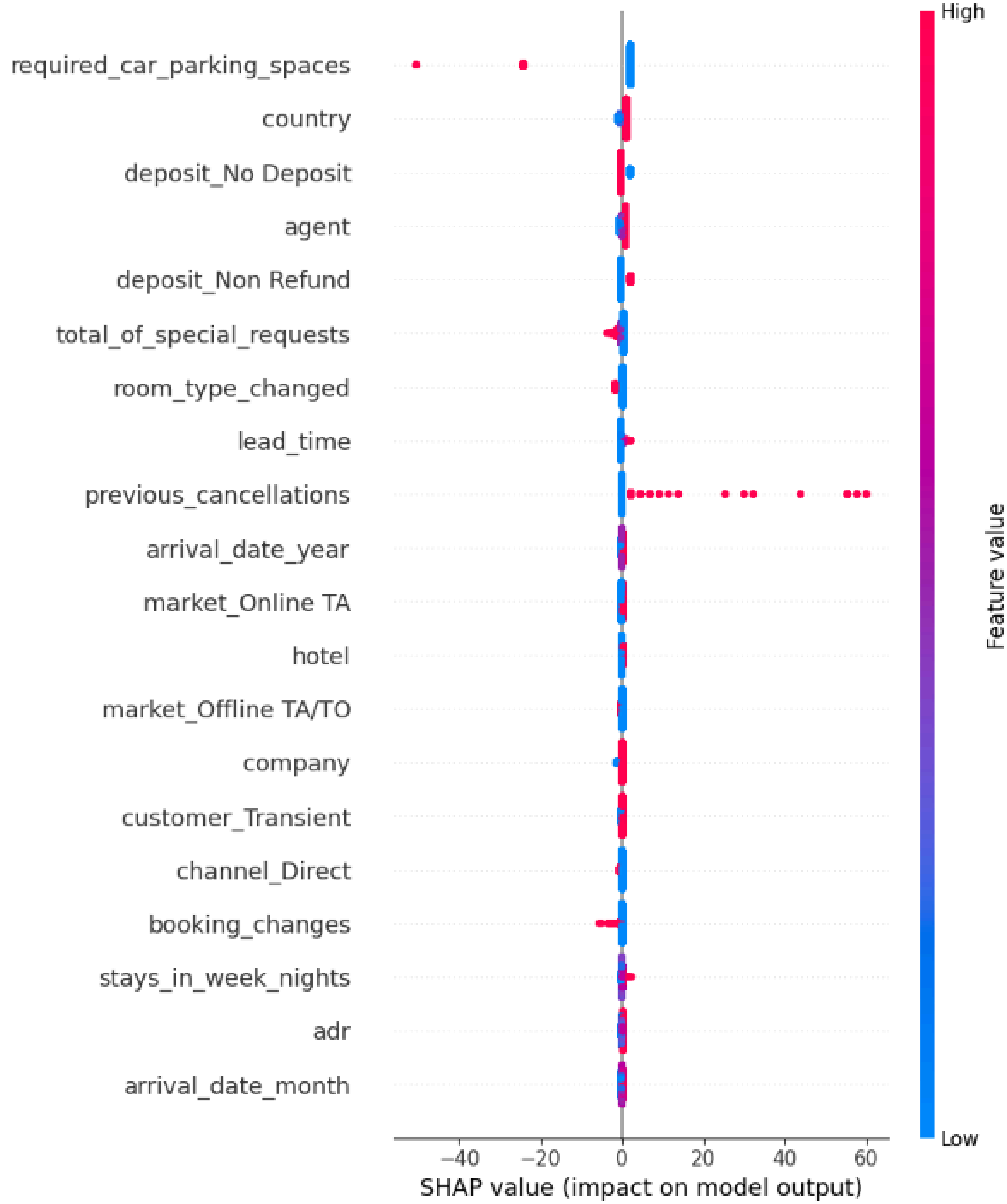
2. Geographical and Booking Characteristics

- country: This feature demonstrates variability across regions, with certain countries showing higher or lower tendencies toward cancellations.
- agent: The booking agent has a modest positive influence, potentially reflecting specific agents' association with certain customer behaviors or demographics.

4c) SHAP

SHAP is a game-theory-based method that measures how each feature contributes to the final prediction. Features with positive SHAP values increase the probability of cancellation, while those with negative SHAP values decrease it, meaning they reduce the likelihood of cancellation. The magnitude of the SHAP value represents how strongly the feature impacts the model's prediction—whether positively or negatively.

We generated SHAP summary plots to visualize how each individual feature affects the probability of cancellation. This visual representation allowed us to clearly see which features were driving cancellations and which were helping reduce them.The summary plot is shown below:


required_car_parking_spaces
country
deposit_No Deposit
agent
deposit_Non Refund
total_of_special_requests
room_type_changed
lead_time
previous_cancellations
arrival_date_year
market_Online TA
hotel
market_Offline TA/TO
company
customer_Transient
channel_Direct
booking_changes
stays_in_week_nights
adr
arrival_date_month
−40 −20 0 20 40 60
SHAP value (impact on model output)
High
Feature value
Low

The SHAP summary plot above demonstrates how different features influence the likelihood of cancellation, ordered by their overall impact. The plot reveals several key insights into the model's decision-making process:

1. Global Feature Importance:

- The feature required_car_parking_spaces appears as the most influential predictor. Lower values of this feature (in blue) reduce the likelihood of cancellation, likely because parking requests indicate a stronger travel commitment. Conversely, higher values of this feature (in red) can occasionally increase cancellation probability, perhaps reflecting overbooking or unfulfilled expectations.
- country also has a substantial impact on cancellations. The broad range of SHAP values associated with this feature suggests significant variability in cancellation behavior across different countries. Certain countries (red values) are associated with a higher likelihood of cancellations, while others (blue) show reduced cancellations.

2. Influence of Deposits:

- The feature deposit_No_Deposit has a predominantly negative impact (blue) on cancellation likelihood. Bookings without a deposit are associated with fewer cancellations, reflecting higher customer confidence or stricter financial commitments.
- Conversely, deposit_Non_Refund contributes positively (red) to cancellation probability. This finding is counterintuitive, as non-refundable deposits typically discourage cancellations. It may indicate that customers choosing non-refundable deposits are often forced to cancel despite the penalty, suggesting external pressures or over-optimistic planning.

3. Booking History and Flexibility:

- previous_cancellations is another highly impactful feature. Higher values (in red) are strongly associated with an increased likelihood of cancellation, as customers with a history of cancellations are more prone to repeating this behavior.
- On the other hand, previous_bookings_not_canceled reduces cancellation risk (blue), indicating that a positive booking history predicts more reliable commitments.

4. Special Requests and Lead Time:

- The feature total_of_special_requests displays mixed impacts. Higher values (blue) typically reduce cancellation risk, as customers who make specific requests may be more invested in their stay. However, in certain cases (red), these requests might signal higher expectations, which, if unmet, could lead to cancellations.
- lead_time has a similar dual impact. Long lead times (red) often increase the likelihood of cancellations, as customers have more time to reconsider their plans. Shorter lead times (blue), however, reduce cancellations by reflecting more definitive planning.

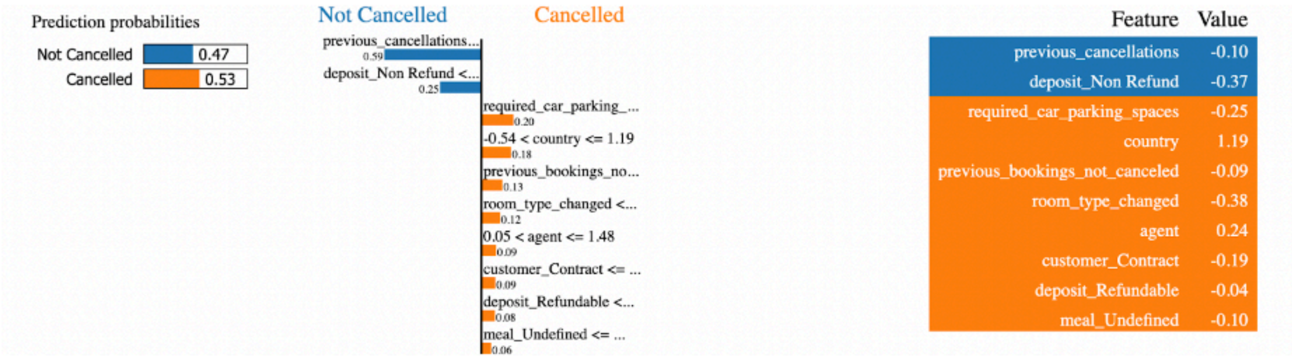
5. ADR (Average Daily Rate):

- The feature adr (average daily rate) displays variability in its impact. Higher values (red) slightly increase cancellation probabilities, potentially due to financial concerns or customers finding cheaper alternatives. Conversely, lower ADR values (blue) reduce cancellations, reflecting affordability or less risk of cost-driven cancellations.

4d) LIME

LIME is an explainable AI tool that helps explain how a complex machine learning model makes its predictions for a specific instance. Unlike SHAP, which shows the impact of each feature across the entire dataset, LIME focuses on explaining a single prediction in detail. It does this by creating a simpler version of the model for that particular data point, like a basic linear model.

In our project, we used LIME to understand why the model predicted that one specific booking would be cancelled. This helped us identify which features were most important for that prediction, making the model's behaviour easier to understand for that individual case. The LIME visualisation showed two sets of features: those that supported the cancellation (in blue) and those that supported not cancelling (in orange).



The LIME chart for the first instance of the training dataset predicted a 53% chance of cancelation and 47% chance of the converse, providing us with the following insights:

1. Key Features Driving Cancellation:

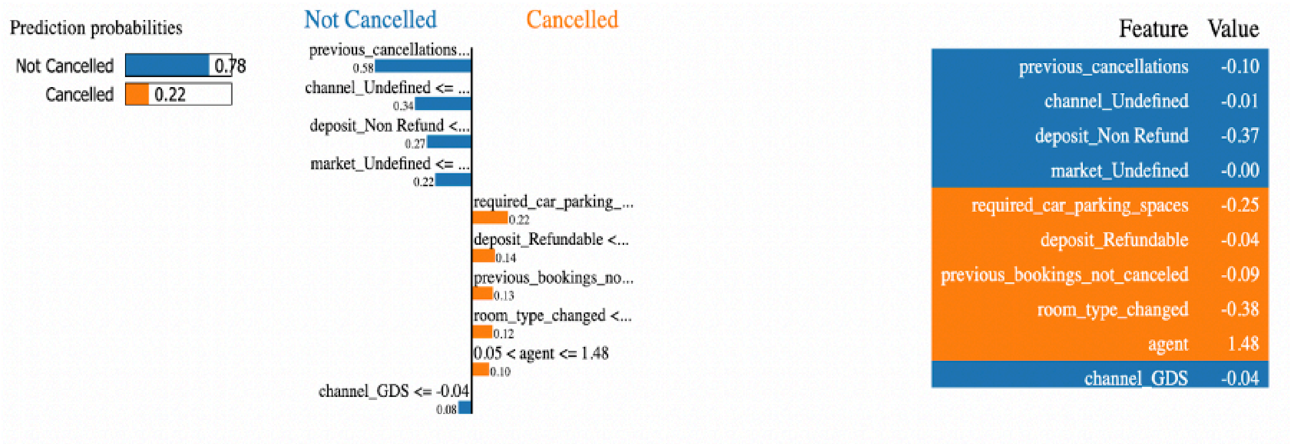
- previous_cancellations (-0.59): This feature significantly contributes to the cancellation probability, as the individual had a history of cancellations. A negative contribution here indicates that the number of previous cancellations aligns with the likelihood of the booking being canceled.
- deposit_Non_Refund (-0.37): Non-refundable deposits are the second most significant driver of cancellations, strongly supporting the likelihood of the customer canceling their booking. This aligns with trends observed in the SHAP analysis, where customers opting for non-refundable deposits are sometimes compelled to cancel despite penalties.

2. Key Features Reducing Cancellation Likelihood:

- required_car_parking_spaces (-0.25): The individual's parking request indicates a stronger commitment to traveling, thereby lowering the likelihood of cancellation.
- country (1.19): Specific characteristics tied to the customer's country may also reduce the likelihood of cancellation, with this feature balancing other cancellation-driving factors.

3. Additional Observations:

- previous_bookings_not_canceled (-0.09): A history of reliable bookings further lowers the probability of cancellation, reinforcing the role of positive booking behavior in predicting non-cancellations.
- room_type_changed (-0.38): A room type change contributes slightly towards reducing cancellation likelihood. This may suggest the customer adjusted their plans rather than canceling outright.



The LIME chart for the first instance of the test dataset predicted as 22% probability of cancelation and a 78% probability of the opposite, shedding light on the following insights:

1. Key Features Driving Cancellation:

- channel_Undefined (-0.34): The channel of booking plays a substantial role in supporting the prediction of cancellation. Undefined channels can indicate higher cancellation rates due to less structured booking processes or lack of direct customer interactions.
- deposit_Non_Refund (-0.27): Again, non-refundable deposits drive the likelihood of cancellation for this test instance, consistent with its impact in the training data.
- market_Undefined (-0.22): An undefined market channel contributes to the cancellation probability.

2. Key Features Reducing Cancellation Likelihood:

- required_car_parking_spaces (-0.25): As observed in the training data, the request for parking indicates a stronger commitment to the booking, reducing cancellation likelihood significantly.

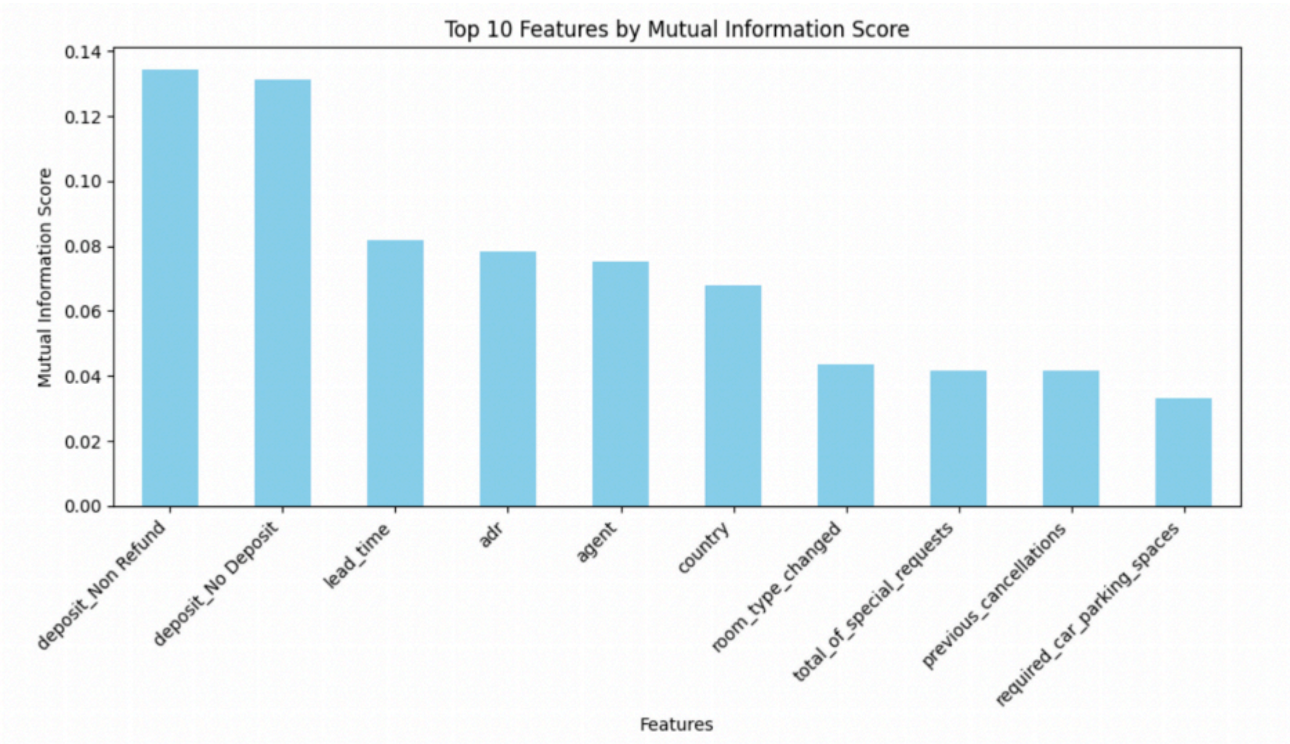
- deposit_Refundable (-0.04): A refundable deposit reduces the likelihood of cancellation slightly.

3. Additional Observations:

- room_type_changed (-0.38): A room type change continues to reduce cancellation probability, as it might indicate successful negotiations or adjustments by the customer to avoid canceling.
- agent (1.48): The booking agent plays a critical role in reducing cancellations for this instance, likely by providing better communication or follow-up to ensure the customer proceeds with the booking.

4e) Mutual Information

Mutual Information (MI) is a feature utility metric that quantifies the association between the features and the target variable. It can capture any kind of relationship, whether linear or non-linear, making it very advantageous for our model. The MI scores for the top 10 features in our dataset are visualized in the bar plot below.



1.Dominant Features

- deposit_Non_Refund (MI ≈ 0.14): This feature exhibited the highest MI score, indicating that non-refundable deposits have a significant predictive relationship with cancellations. Customers who chose non-refundable deposits were more likely to cancel.
- deposit_No_Deposit (MI ≈ 0.14): Bookings without deposits were also highly informative. Unlike non-refundable deposits, this feature negatively correlates with cancellations.

2. Other High-Impact Features

- lead_time (MI ≈ 0.08): Longer lead times are associated with a higher likelihood of cancellations, as customers have more time to reconsider or face changes in circumstances.
- adr (Average Daily Rate, MI ≈ 0.08): Higher booking costs correlate with cancellations, potentially reflecting customer sensitivity to price changes.
- agent (MI ≈ 0.07): The booking agent impacts cancellations, with specific agents likely tied to customer demographics or booking processes that influence behavior.

3. Commitment Indicators

- total_of_special_requests (MI ≈ 0.06): Customers who make special requests are less likely to cancel, reflecting their investment in the booking process.
- required_car_parking_spaces (MI ≈ 0.05): Parking space requests significantly reduce the likelihood of cancellation, indicating a higher level of commitment to travel.

Algorithm 5: CatBoost

5a) Classification report, Accuracy, AUC

```
Test Set Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.92      0.91     14951
           1       0.86      0.82      0.84      8847

    accuracy                           0.88     23798
   macro avg       0.88      0.87      0.87     23798
weighted avg       0.88      0.88      0.88     23798
```

	Dataset	Accuracy	AUC
0	Train	0.893738	0.963548
1	Test	0.883856	0.955943

As shown in the tables above, the training dataset achieved an accuracy of 89.4% and an AUC of 96.4%, while the test dataset had slightly lower metrics, with an accuracy of 88.4% and an AUC of 95.6%. This indicates a well-generalized model that performs consistently across both training and testing datasets, with minimal signs of overfitting. The high AUC means that the model discerns efficiently between the two classes (class 0 and class 1).

```
Test Set Classification Report (Optimized Model):
              precision    recall  f1-score   support

           0       0.90      0.92      0.91     14951
           1       0.87      0.83      0.85      8847

    accuracy                           0.89     23798
   macro avg       0.88      0.88      0.88     23798
weighted avg       0.89      0.89      0.89     23798
```

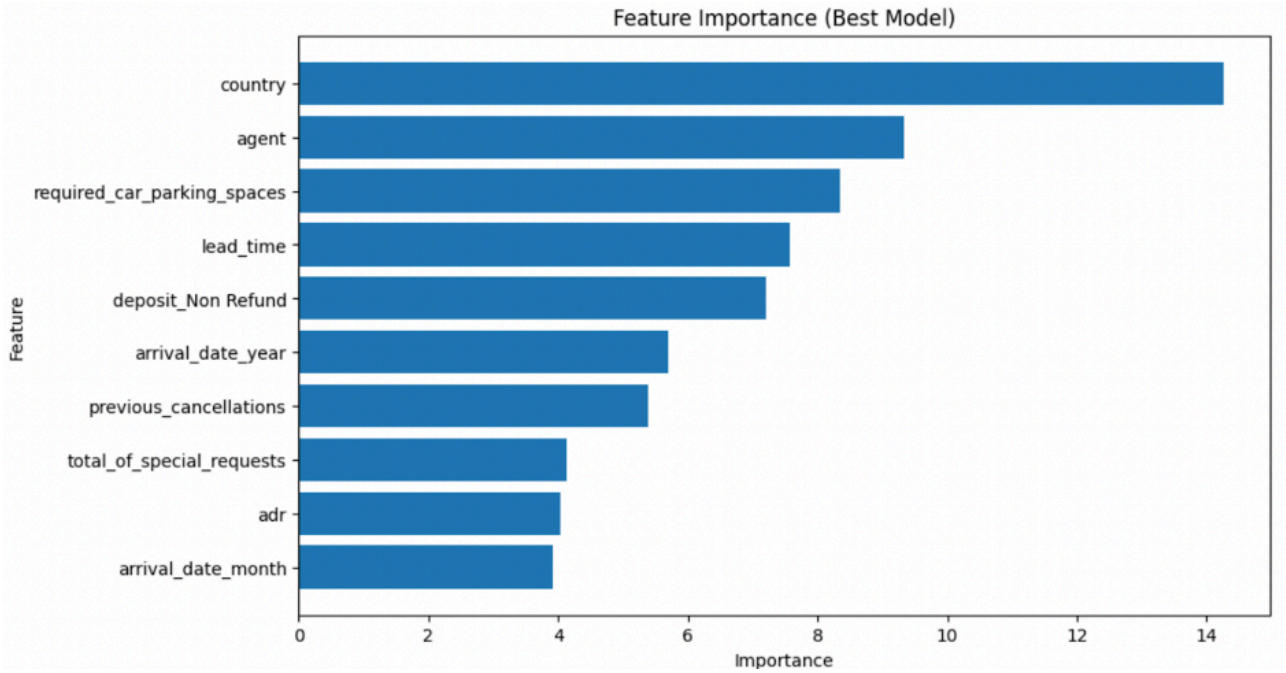
```
Complete Performance Summary:
```

	Dataset	Accuracy	AUC
0	Train	0.893738	0.963548
1	Test	0.883856	0.955943
2	Train (Optimized)	0.917743	0.978011
3	Test (Optimized)	0.889487	0.959580

Following optimization, there were minor improvements in accuracy, with training accuracy increasing from 89.4% to 91.8% and the testing datasets demonstrating negligible improvement from 88.4% to 88.9%. The AUC also showed very modest gains within a percent of change. This shows that the original configuration of the baseline model was already close to optimal.

5b) Feature Importance

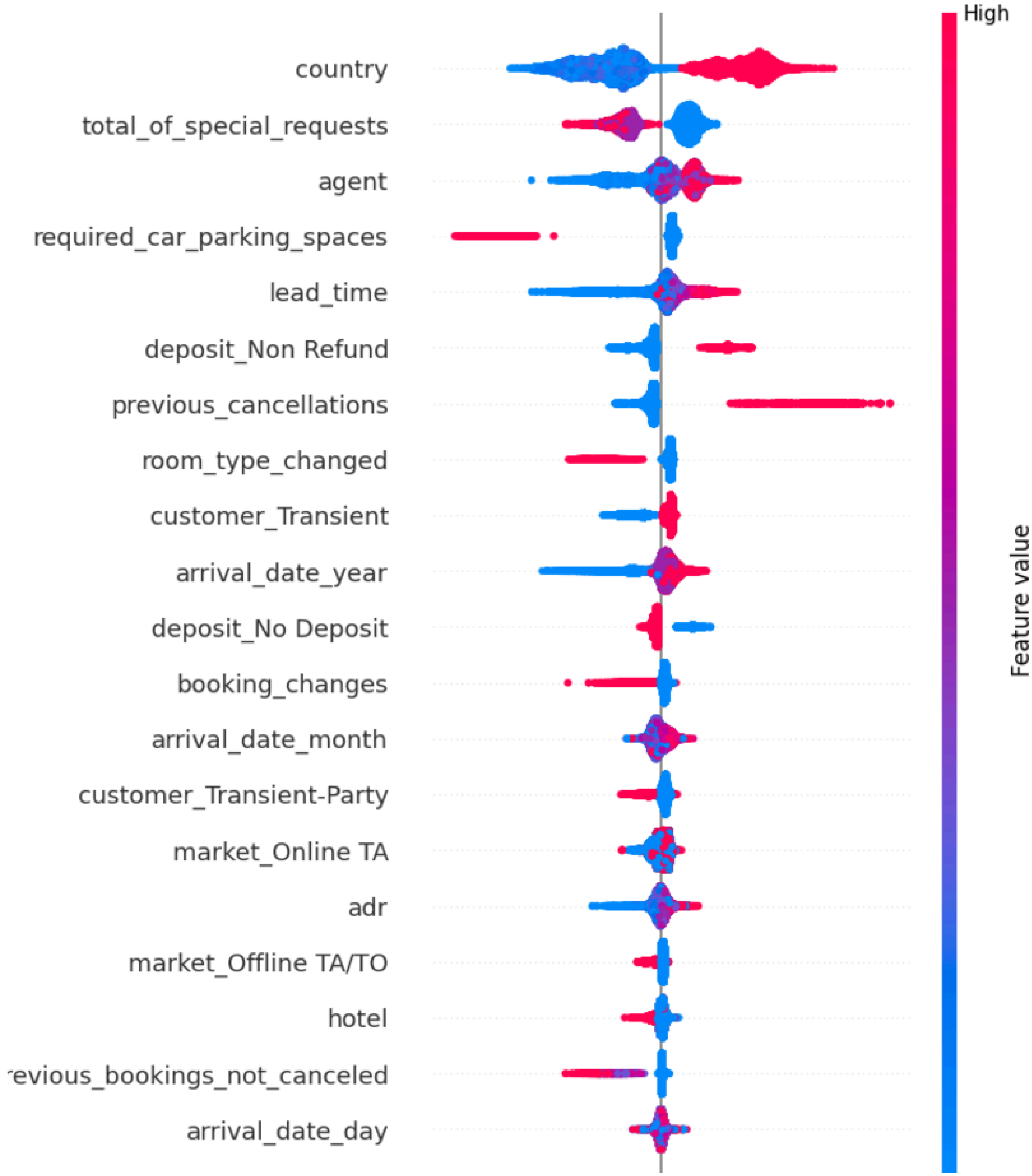
Feature importance in CatBoost works by ordering features that are the most important in making correct predictions. If the omission of the feature significantly alters the model's performance, it is deemed significant.

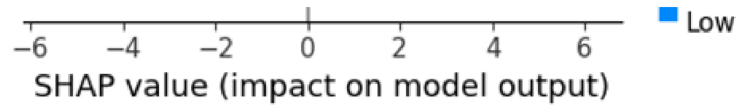


The feature importance plot above demonstrates the most critical predictors influencing the CatBoost model's classification. The "country" feature was the most significant, indicating that the geographical origin of the bookings greatly affects the likelihood of cancellations. The "agent" feature secured the second highest rank, suggesting that the use of a booking intermediary played a crucial role. "Required car parking spaces" was also a key factor, signifying higher investment in the booking.

Other influential variables include "lead_time," which reflects the booking window before the intended stay. A longer lead time may provide customers with more opportunities to cancel, hence its importance. "Deposit_Non_Refund" also holds substantial weight, confirming that the non-refundable nature of deposits discourages cancellations.

5c) SHAP


country
total_of_special_requests
agent
required_car_parking_spaces
lead_time
deposit_Non Refund
previous_cancellations
room_type_changed
customer_Transient
arrival_date_year
deposit_No Deposit
booking_changes
arrival_date_month
customer_Transient-Party
market_Online TA
adr
market_Offline TA/TO
hotel
revious_bookings_not_canceled
arrival_date_day
High
Feature value

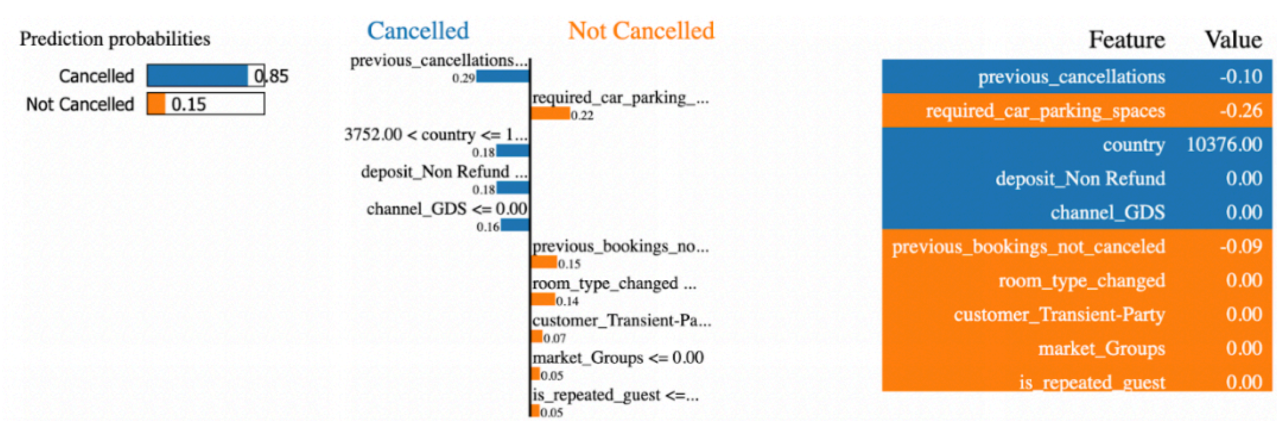


The SHAP plot illustrates that " country" was the most dominant feature, with high SHAP values for specific countries significantly affecting the model's output. This suggests that customer origin plays a critical role in determining cancellation probabilities, likely reflecting regional booking behaviors.

"Total_of_special_requests" has a notable impact, with higher values generally associated with reduced cancellation likelihood, indicating a stronger commitment from customers who make specific arrangements. Similarly, "Agent" and "Required_car_parking_spaces" show considerable influence. Certain booking agents correlate with higher or lower cancellation tendencies, while parking space requirements may indicate a more engaged visit with personal travel commitments, reducing cancellations.

"Lead_time" demonstrates its typical pattern—longer lead times generally increase cancellation risk, as customers with extended timelines have more opportunities to change plans. "Deposit_Non_Refund" reduces the likelihood of cancellations due to the financial loss associated with canceling a non-refundable deposit.

5d) LIME

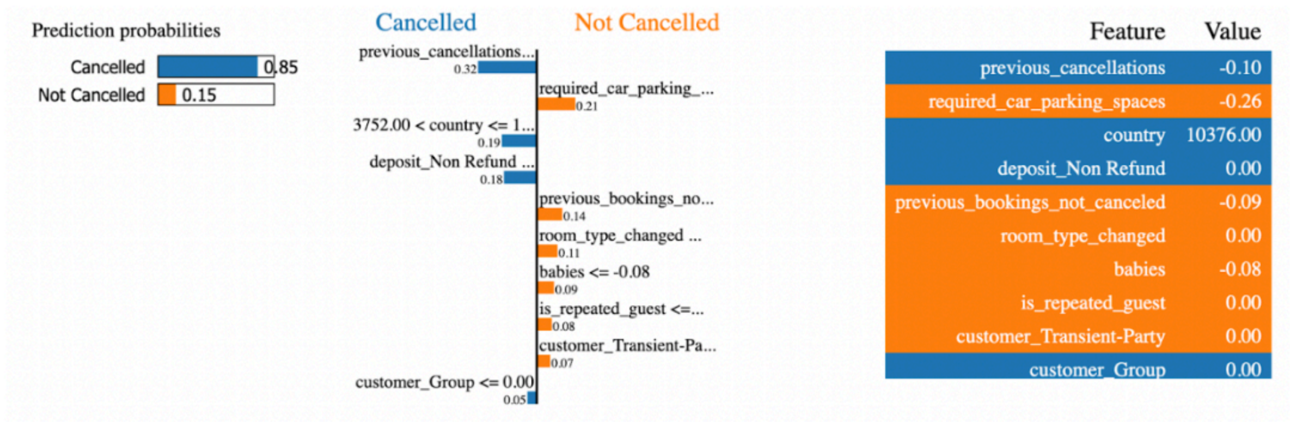


This LIME visualization for the first row of the training dataset illustrates that the probability of this booking had an 85% chance of being canceled.

For the "Cancelled" prediction, previous_cancellations held the most weight, with customers having a history of cancellations being strongly correlated with future cancellations, reflecting customer reliability.

The country feature plays a significant role in the "Cancelled" prediction, suggesting that the geographical origin of the customer affected the likelihood of cancelation. Deposit_Non_Refund contributes notably to cancellations, as customers who opt for non-refundable deposits are less likely to cancel due to the financial implications.

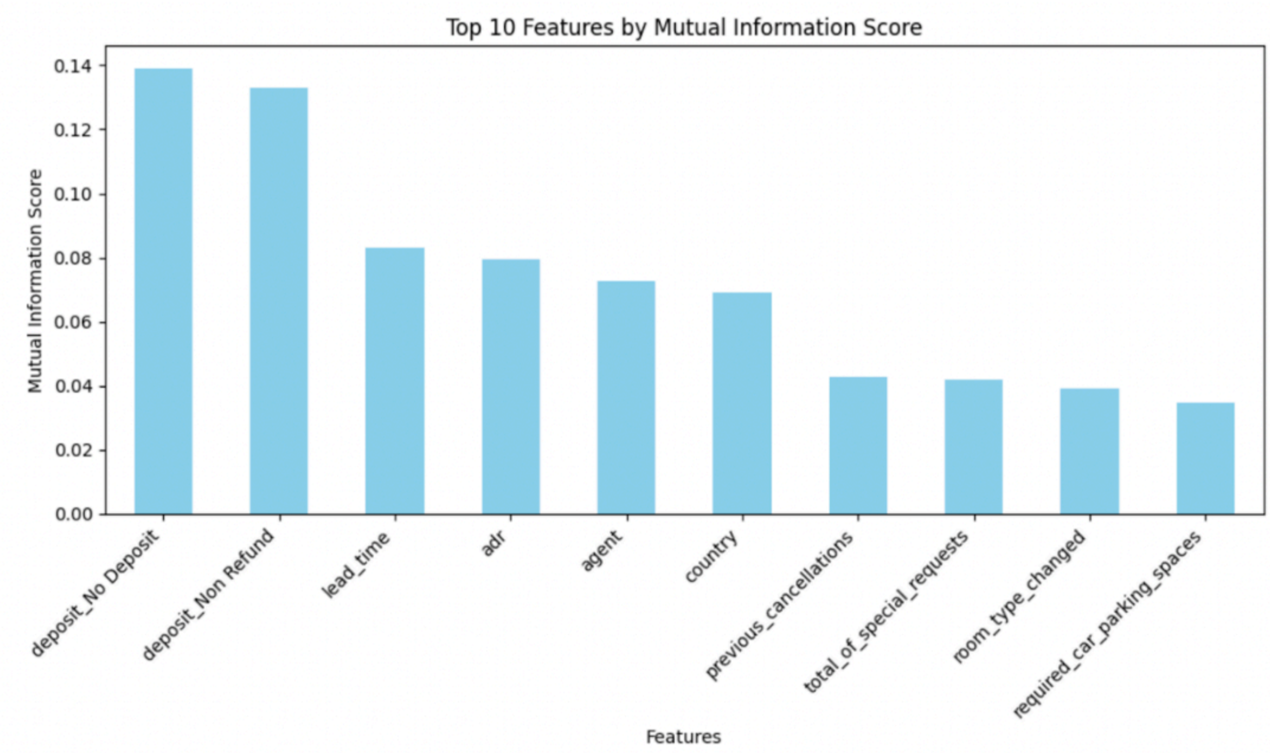
On the "Not Cancelled" side, required_car_parking_spaces and previous_bookings_not_canceled emerge as critical factors. For the former, the requirement of parking spaces indicates a higher investment in the hotel stay due to a physical trip and for the latter, customers who have a history of non-cancelled bookings are generally more committed to their reservations, explaining their behavior in the plot.



For the first instance of the training data, the LIME visualization for the test data depicts an 85%change of cancelation. For the "Cancelled" class, “previous_cancellations” is the most significant feature, strongly indicating that a history of cancellations increases the chances of the current booking being cancelled. Similarly, “country” and “deposit_Non_Refund” play an influential role, indicating regional differences and the impact of financial commitments on the likelihood of cancellation.

On the other hand, for the "Not Cancelled" class, “required_car_parking_spaces” was the dominant indicator, emphasizing that customers with physical travel plans are less likely to cancel. Additionally, previous_bookings_not_canceled and room_type_changed also support the "Not Cancelled" prediction, reflecting the influence of consistent past behavior and booking adjustments.

5e) Mutual Information



The mutual information plot highlights the most influential features based on their predictive strength for the target variable. The feature "deposit_No_Deposit" ranks highest, indicating its strong impact on predicting cancellations. This suggests that bookings without deposits are far more likely to be canceled compared to those with deposits. Similarly, "deposit_Non_Refund" closely follows, reinforcing the observation that financial commitment significantly reduces cancellation tendencies.

"Lead_time" is another critical feature, with its high mutual information score implying that the longer the time between booking and check-in, the higher the likelihood of cancellations. This is expected as extended lead times provide greater flexibility for customers to modify their plans. "Adr" (average daily rate) also shows substantial influence, with higher rates potentially deterring cancellations due to increased financial stakes.

"Agent" and "country" emerge as significant contributors, suggesting that booking agents and customer demographics play important roles in determining cancellation behavior. For example, specific agents or regions might be associated with higher or lower cancellation rates based on their customer base or booking policies.

Features like "previous_cancellations" and "total_of_special_requests" are moderately influential. A history of cancellations increases the probability of future cancellations, whereas a higher number of special requests often correlates with lower cancellation likelihood, reflecting greater customer commitment.

Lastly, "room_type_changed" and "required_car_parking_spaces" exhibit lesser but notable contributions. Changes in room type might indicate customer uncertainty, while parking requirements could signify planned commitments, reducing cancellations. In summary, these insights provide a clear understanding of the factors driving cancellations, enabling targeted strategies to mitigate them effectively.

Algorithm Comparisons

The five algorithms: XGBoost, MLP, LightGBM, Logistic Regression, and CatBoost all demonstrate individual strengths and weaknesses which are quantified for us through various previously discussed performance metrics that provide valuable key insights used for comparison.

a. Performance and Generalization:

CatBoost and LightGBM seemingly had the highest accuracy (at 88.89% and 88.9%) and AUC (95.5% and 95.6% respectively), which demonstrated their ability to effectively model complex relationships whilst simultaneously maintaining good generalizations towards unseen data. Although, it seems as if logistic regression and MLP tended towards exhibiting only more moderate performance, which indicated that for these algorithms, there was still potential for improvement. While logistic regress ion is consistent and robust, it lagged in accuracy and AUC which spoke towards this algorithm's inability to accurately capture nonlinear relationships effectively.

b. Feature interpretability through explainable AI:

Tree-based models such as CatBoost and XGBoost seemed to demonstrate a distinct advantage in generating actionable insights through SHAP and LIME visualisations. These techniques were able to highlight both local as well as global feature importance, and overall had the ability to enhance model trust and usability. MLP, despite being less inherently interpretable, leveraged SHAP to provide insights, although it lacked the depth of tree-based models. Tree-based models, on the other hand, seem to be particularly effective at generating such insights due to a hierarchical structure as well as their strong ability of capturing nonlinear interactions and relationships. These models split features into decision thresholds, which allow us to easily trace back to how a specific prediction is linked to specific feature contributions. This aligns well with general explainable AI framework. By leveraging the inherent structure of decision trees, tree-based models provide a higher interpretability that SHAP and LIME can isolate and quantify to provide more nuanced insights compared to models that assume linearity like Logistic Regression.

c. Optimization and Scalability

Tree-based algorithms tend to show a stronger performance and scalability, particularly after hyperparameter tuning. Logistic Regression and MLP require minimal tuning, which makes them easier to implement but with a trade-off in predictive power.

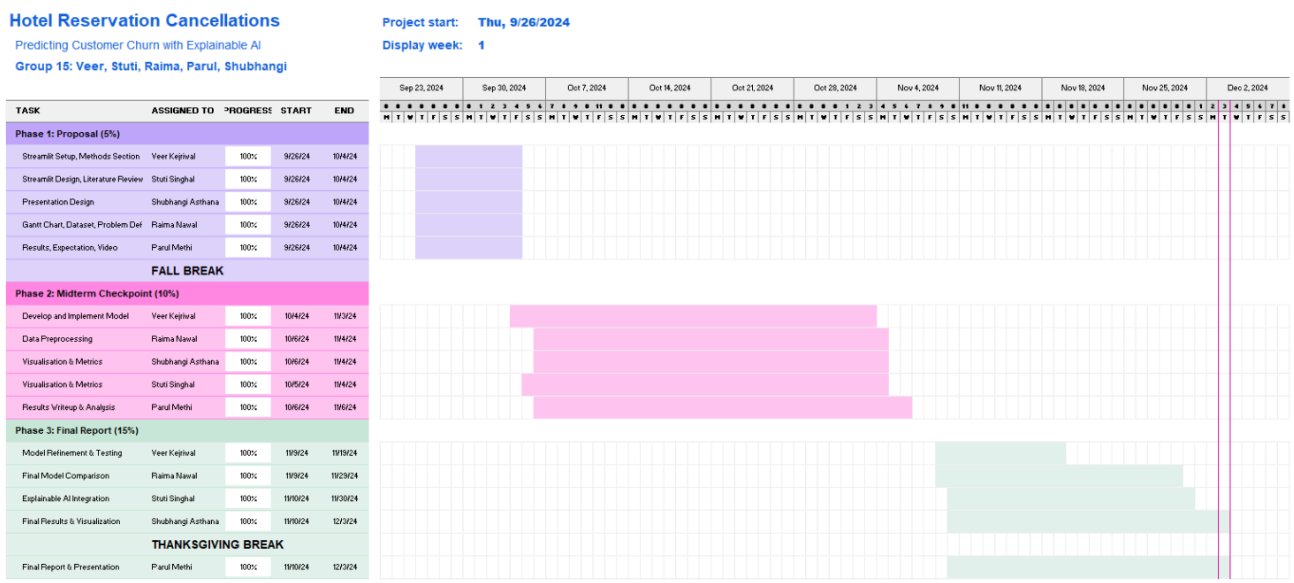
To summarize, CatBoost and XGBoost emerge as top performers in overall terms of accuracy, AUC, and feature interpretability. It seems that for high-performance requirements with actionable insights, tree-based models are more preferable, while simpler models such as Logistic Regression are better suited for less complex scenarios.

Next Steps

While in the terms of this course, our project is final and complete, there are still various next steps worth considering that focus on refining the models further to better leverage insights and improve the overall machine learning decision-making process. First, through a comparison of the models done earlier, it is apparent that there are some models working with a better cumulative performance over others. It may be worth considering further tuning such top-performing models to optimize their performance more with the knowledge now that they have demonstrated a higher rate of success. Techniques like Bayesian Optimization or further automated hyperparameter tuning could be employed to help find the best configurations efficiently. Additionally, it may be worth incorporating domain-specific knowledge into feature engineering - such as creating interaction terms or normalizing features to potentially help uncover new patterns and improve predictions overall. Ensemble techniques that create hybrid models is another area open to exploration that could help potentially boost overall accuracy and robustness.

From an application perspective, the insights gained through explainable AI can be translated into real actionable insights. For example, the strong influence of “depost type” on cancellations suggest a need for adjusting cancellation policies to balance flexibility and commitment. Features such as “lead_time” and “special_requests” can also guide targeted interventions - this can include options such as sendin reminders or offering consumer incentives for early cancellations to reduce last-minute changes. Deploying these models in production environments with intermitten retraining can help it stay current and adaptable according to consumer trends to ensure sustainability and reliability.

Project Gantt Chart:



Project Contribution Table

	Name	Contribution
0	Veer Kejriwal	Model initialization coding, model optimization coding, GitHub management and organization, interpretation of models
1	Stuti Singhal	Explainable AI (shap, lime) coding, model optimization coding, interpretation of models and vizualizations
2	Raima Nawal	Explainable AI (shap, lime) coding, model optimization coding, interpretation of models and vizualizations
3	Shubhangi Asthana	Feature importance coding, interpretation of vizualizations
4	Parul Methi	Mutual information coding, coding the streamlit website, recording the video

References

1. Antonio, N., Almeida, A. M., & Nunes, L. (2017). Predicting hotel booking cancellations to decrease uncertainty and increase revenue. *Tourism & Management Studies, 13*(2), 25–39. https://doi.org/10.18089/tms.2017.13203

2. Lin, Y. (2023). Research on the influencing factors of cancellation of hotel reservations. *Highlights in Science, Engineering and Technology, IFMPT 2023, 61*, 107–118. https://doi.org/10.18089/rs.IFMPT2023-107118

3. Sánchez, E. C., Sánchez-Medina, A. J., & Pellejero, M. (2020). Identifying critical hotel cancellations using artificial intelligence. *Tourism Management Perspectives, 35*, Article 100718. https://doi.org/10.1016/j.tmp.2020.100718

4. Özkurt, C. (2024). Transparency in decision-making: The role of explainable AI (XAI) in customer churn analysis. *Sakarya University of Applied Sciences*. https://doi.org/10.21203/rs.3.rs-3937355/v1