Information Technology & Tourism (2026) 28:9
https://doi.org/10.1007/s40558-025-00349-9

ORIGINAL RESEARCH

Navigating uncertainty: enhancing hotel cancellation predictions with adaptive machine learning

Pedro Silvestre[1] · **Nuno Antonio**[1,2] · **Paulo Carrasco**[2,3]

Received: 30 May 2025 / Revised: 18 October 2025 / Accepted: 28 November 2025

Abstract

Accurately predicting hotel booking cancellations is critical for hotel management, especially during volatile periods such as the COVID-19 pandemic. Prior work demonstrated that machine-learning (ML) models perform well on historical data, yet few studies test robustness under severe disruption. We evaluate ML classifiers trained on pre-pandemic data from four hotels and assess their adaptability to pandemic conditions (Study One). We then examine whether adding pandemic observations via a dynamic sliding-window approach improves accuracy (Study Two). Pre-pandemic models exhibit reasonable discrimination, but including pandemic-period data can raise the Area Under the Curve (AUC) by up to 5% points. A nine-month training window balances stability and responsiveness, capturing rapid shifts in booking patterns and customer behavior. Feature importance also changes: Lead time and other drivers show altered effects during the pandemic, underscoring the need for continuously updated models. Anchored in concept-drift theory, we interpret the pandemic as an abrupt shift in the cancellation decision boundary and show that sliding-window retraining together with interpretable diagnostics (e.g., the Lead time crossover threshold) provides a theoretically grounded blueprint for prediction under distributional change. Our results advocate scheduled retraining and lightweight drift diagnostics to sustain forecast accuracy and managerial actionability. For hotel managers and technology providers, the proposed approach supports proactive cancellation management, more reliable forecasting, and resilient operations in volatile markets, demonstrating the robustness of adaptive ML under conditions of extreme market volatility. The study advances theoretical understanding and practical applications by operationalizing concept-drift management in revenue-critical settings.

Keywords Concept drift · Crisis · Data science · Hospitality · Machine learning · Predictive modeling · Booking cancellations

---

Extended author information available on the last page of the article

1 Introduction

Accurately predicting hotel booking cancellations is a significant challenge in the hospitality industry, particularly during periods of high volatility. Accurate cancellation prediction enhances demand forecasting and supports targeted managerial interventions (António 2019a). The onset of the 2019-nCoV (COVID-19) pandemic, with initial cases reported in December 2019, its declaration as a global pandemic by the World Health Organization on March 11, 2020 (WHO 2020), and major disruptions peaking in 2020–2021 through travel restrictions and variants like Delta, serve as a prime example of such a high-volatility period. The pandemic led to widespread travel restrictions and a decreased willingness to travel, profoundly impacting the hospitality sector with massive cancellations and significantly affecting hotel performance (António and Rita 2020). Note that no-shows are treated as cancellations in this study, consistent with prior research (António et al. 2017b).

In economic and financial contexts, "high-volatility times" refer to periods characterized by considerable uncertainty and price fluctuations, often triggered by political events, economic data releases, corporate earnings announcements, or financial results, or natural disasters (Schwert 1989). While not always detrimental, events like pandemics invariably introduce negative volatility. Before the COVID-19 crisis, research had demonstrated the feasibility of developing predictive models for booking cancellations using historical booking data, such as XGBoost and Random Forest classifiers. However, these existing models were not typically designed or tested for extreme conditions like a pandemic, where mass cancellations become commonplace. Consequently, a critical research gap exists in understanding how pre-pandemic cancellation models perform under such duress and how their predictive capabilities can be enhanced or adapted to maintain accuracy during and after such crises. This research aims to address this gap by first validating whether Machine Learning (ML) models trained on pre-COVID-19 data can still effectively predict booking cancellations during the pandemic. Secondly, it seeks to identify changes in cancellation-driving features and customer behavior and explore how models can be improved by incorporating more recent data, including data from the pandemic period itself.

Hotel booking cancellations can stem from factors beyond a customer's control, such as alterations in travel plans, illness, accidents, or adverse weather conditions (Chen et al. 2011; Falk and Vieru 2018). Alternatively, cancellations may result from deliberate customer actions, including finding a hotel with a better price, a more desirable location, superior services or facilities, or deciding to relocate to be near friends or family (Chen et al. 2011). Generally, an accurate forecast of whether a customer will cancel their booking is crucial for effective inventory allocation, pricing strategies (Mehrotra et al. 2006), and other managerial decisions related to staffing, supply procurement, and overall profitability or cash flow management (Hayes and Miller 2011). The financial implications of cancellations are significant, motivating research into profit-driven prediction models, as explored by Liu et al. (2024), who developed approaches to address cancellation prediction by directly incorporating profitability outcomes.

Previous research has shown that it is possible to achieve strong predictive performance in the cancellation of bookings in the hospitality industry. Equipped with cancellation prediction models that can estimate booking cancellation outcomes with high accuracy, hotels can proactively contact customers identified as likely to cancel and take preventative actions (António 2019a; António et al. 2017b, 2019c). Understanding the evolution of booking cancellation patterns, especially during high-volatility periods like the COVID-19 pandemic, offers actionable insights for hoteliers. This knowledge is vital for adapting booking policies and refining demand forecasting strategies in unpredictable environments, thus underscoring the practical importance of this research for hotel managers navigating such challenging times.

Hotel booking cancellations are both frequent and economically material. Prior studies report that cancellations account for roughly one-fifth of bookings and may exceed 60% in airport or roadside contexts, with city properties often exhibiting higher cancellation rates than resort properties (Romero Morales and Wang 2010; António 2019a). Because cancellations erode forecast accuracy and revenue, accurate prediction supports inventory control and price discrimination, informs staffing and procurement decisions, and, ultimately, cash-flow management (Mehrotra et al. 2006; Hayes and Miller 2011). Within this managerial context, our study examines how predictive performance and the salience of key drivers change under high volatility.

This work aims to demonstrate the application of data science in revenue management for predicting hotel booking cancellations during periods of high volatility. It utilizes uncensored data from the Property Management Systems (PMS) of two resort hotels in the Algarve region and two city hotels in the Lisbon region of Portugal. To the best of our knowledge, limited research has focused explicitly on testing and improving booking cancellation forecast models using COVID-19 data, thereby extending the existing body of knowledge. This study intends to fill this void by conducting two distinct analyses: the first involves creating and testing a classification ML model trained solely on pre-COVID-19 data against bookings made during the COVID-19 period. The second study focuses on developing additional ML models trained on more recent data, including data from the pandemic's onset, using different sampling and a sliding-window training approach. The findings from both studies demonstrate that while pre-pandemic models remain somewhat effective, incorporating pandemic data using a sliding window approach significantly improves predictive accuracy.

2 Literature review

2.1 Machine learning on booking cancellation prediction

With the emergence of ML, its application in the hotel industry rapidly grew. Zhang (2019) applied multiple classical ML models in hotel demand forecasts based on pricing location, concluding that these approaches beat the traditional models. Researchers like António (2019a) have extensively analyzed past work on this topic, concluding that there is a tendency to employ detailed booking data and advanced

ML classification algorithms. More recently, Herrera et al. (2024) and Yoo et al. (2024) also explored machine learning approaches for forecasting hotel cancellations, further highlighting the ongoing development and application of these techniques in the industry.

Previous studies using a similar data structure and hotel types showed promising results. For resort hotels, using hotel PMS data ranging from January 1, 2016, to November 20, 2017, the models reached an Area Under the Curve (AUC) between 0.651 and 0.829. For city hotels, the AUC ranged from 0.773 to 0.864. The same author experimented with a shorter timeframe, from August 1, 2016, to November 20, 2017, and achieved an AUC from 0.644 to 0.818 for resort hotels. For city hotels, the AUC ranged from 0.771 to 0.923 (António 2019a).

However, previous studies did not explicitly account for temporal distribution shift (concept drift)—i.e., changes in the joint data-generating process that affect either the predictor distribution P(x) ("virtual drift") or, more critically, the conditional mapping P(y|x) ("real drift"), particularly salient under crisis conditions. This omission limits external validity when models are exposed to regimes that differ from their training data (Žliobaitė et al. 2016).

Predictive ML models such as XGBoost benefit from post-hoc interpretability via SHAP (Shapley Additive Explanations), which attributes local and global feature contributions and supports model diagnostics and stakeholder buy-in (Lundberg and Lee 2017). To our knowledge, applications of SHAP to hotel booking cancellation prediction remain scarce; we therefore use SHAP to track how feature salience evolves across pre-pandemic and pandemic regimes. SHAP belongs to the broader field of explainable machine learning (XAI).

Building on this lens, we treat concept drift not merely as context but as a framing device for our empirical strategy: we examine whether the cancellation decision boundary shifts during crises and whether predictor importance rotates across regimes. This strategy motivates a sliding-window training design that privileges recent evidence and an SHAP-based interpretability protocol to monitor time-varying feature importance (Žliobaitė et al. 2016; Baier et al. 2020).

2.2 COVID-19 impact on hospitality

The impact of economic crises on tourism has been studied by several authors (Ritchie et al. 2010; Song and Lin 2010). However, except for new studies involving COVID-19, not many have analyzed the impact of health-related crises on the tourism industry (Novelli et al. 2018). The COVID-19 pandemic, in particular, has caused unprecedented disruption. Among the new hospitality-related COVID-19-centered studies, the focus seems to be on the impact on hotel operations. Specifically, how hotels were affected and what will change in the future in hotel operations (Anguera-Torrell et al. 2021; António and Rita 2020; Hao et al. 2020; Lai and Wong 2020; Vong et al. 2021). More recent post-2021 studies examine recovery strategies (Matijević et al. 2025), lasting influences on accommodation preferences (Ding et al. 2025), and workforce vulnerability (Luo 2025).

One of the measures affecting hospitality was the paralyzation of international travel by airplane. In 2020, tourist trips decreased by 41.1%, reaching 14.4 million

(+ 10.8% in 2019 and + 4.2% in 2018), the lowest value in the last decade. This variation reflects the pandemic experienced from March onwards due to COVID-19, which led to various confinements throughout the year and, consequently, a general slowdown in tourist activity (INE 2021). In 2021, Portugal registered 9.6 million arrivals of non-resident tourists, corresponding to an increase of 48.4% compared to 2020, representing only 39.0% of the value obtained in 2019 (24.6 million) (INE 2022).

3 Methods

This study adopted the well-known Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology (Chapman et al. 2000). The research is structured into two main studies to address the objectives outlined in the introduction.

3.1 Business and data understanding

An Exploratory Data Analysis (EDA) was performed during the Data Understanding phase. The data used contains hotel booking information from four Portuguese hotels – two located in the city of Lisbon (Hotel City 1, C1, and Hotel City 2, C2) and two located in the resort beach area of Algarve (Hotel Resort 1, R1, and Hotel Resort 2, R2), both in Portugal. These four hotels are mid-scale; C1 and C2 are chain-affiliated, benefiting from centralized marketing, while R1 and R2 are independent, relying on local appeal. These were selected to provide a balanced dataset, including city and resort hotels in different regions of Portugal, ensuring diversity in booking patterns and customer behavior. When the data was collected, none of these hotels used any predictive analytical tool nor took proactive action in determining which customers planned to cancel their booking. The datasets contain 670,343 bookings with 30 features (32 when the date feature is decomposed into the day, month, and year). The data were collected on October 1, 2021 (its structure can be seen in Appendix A), with all bookings bounded by this cut-off for consistency. For benchmarking and comparability, the structure of the datasets followed the structure published by (António et al. 2019b). The raw data summary of the four datasets can be seen in Table 1.

Two data processes were applied before proceeding with the EDA. The first process was the removal of bookings that could cause leakage to the models. At any moment in time, bookings with three types of status coexist in a hotel Property Management Systems (PMS) database: (A) Effective – bookings with an arrival date that is before or equal to the current date, for which customers already checked out or are

Table 1 Datasets summary statistics

Hotel	#Rows	Oldest Date	Number of Rows	Number of Columns	Size (MB)
C1	204 567	20/03/2015	204 567	30	31.8
C2	302 698	20/07/2017	302 698	30	48.2
R1	113 177	18/11/2014	113 177	32	15.5
R2	49 901	07/03/2017	49 901	30	7.92

checked in; (B) Cancelled - bookings with an arrival date set for any moment in time (past or future) but which were already canceled; (C) Unknown – bookings with an arrival date after the current date and that have not been canceled. However, the (C) Unknown bookings can still be canceled until the date of arrival. For this study and to achieve model improvement, following past research (António et al. 2017b), all bookings on future dates were removed, i.e., all "C" bookings and "B" bookings with arrival dates in the future. In conformity with what was identified in past research (António et al. 2017a, b), despite the differences between no-shows and cancellations, both will be treated as cancellations. The second process was the application of data consistency rules and filters to remove incorrect data. Because sometimes there are errors committed by the hotel staff during data imputation, standard consistency rules (António 2019a) were applied in all hotels' datasets, namely:

1. Price per night must be null or positive.
2. Number of guests must be positive.
3. Number of nights must be positive.
4. Lead time must be non-negative.

As shown in Fig. 1, more city than resort hotel bookings exist. Hotel C1 seems to have a pronounced seasonality, seen by two higher peaks in the middle of the year and the last quarter. Hotel C2 has grown in the number of bookings from 2018 to the start of the pandemic in 2020; this growth may be attributed to its chain affiliation, which provides robust marketing and distribution networks, unlike the independent resorts. Hotels R1 and R2 had a similar recorded trend, but for R2, it started in March 2017.

There is a pattern of a significant drop in the number of bookings for both hotel types when the COVID-19 pandemic started, followed by a recovery during the summer of 2020. This first drop is particularly pronounced in city hotels C1 and C2. The inflection point is achieved in July 2020 for C2, June for C1, April for R1, and May for R2. Following that recovery, there is another significant drop, starting in 2020 and

Fig. 1 Total bookings per arrival date per hotel

continuing into 2021. Again, there is a recovery in the first quarter of 2021. After this last recovery in 2021, hotels C2 and R1 have more bookings than in the pre-pandemic period, with R2 recovering more slowly and C1 still behind pandemic levels.

On February 29, 2020, the threshold was defined as an identifier for the pandemic's start. The reason was that in March 2020, most European countries officially recognized COVID-19 as a threat to public health and the economy. This threshold was defined using the ArrivalDate feature instead of the ReservationStatusDate, as the ReservationStatusDate could introduce leakage in the non-canceled bookings.

Comparing a timeframe where all hotels have a significant number of bookings and are closer to the pandemic's start, Fig. 2 represents how the cancellation rate evolved from January 2018 onwards, including the start of the COVID-19 pandemic. One can observe that, before COVID-19, hotel C1 had a declining trend of cancellations of around 30% on average, and C2 had a somewhat stable cancellation rate of around 40%. Hotel R1 had around 25% and R2 around 15% cancellation rate. The pre-pandemic results are aligned with previous research (António 2019a). However, once the pandemic started, cancellations quickly rose to unprecedented levels, reaching 100% for several months. Key events, such as the State of Emergency declaration in March 2020 and January 2021, are annotated as vertical lines.

One can observe different months in which hotels have all the bookings canceled. Namely, hotel C1 had 100% cancellations in April, May, and June of 2020 and again

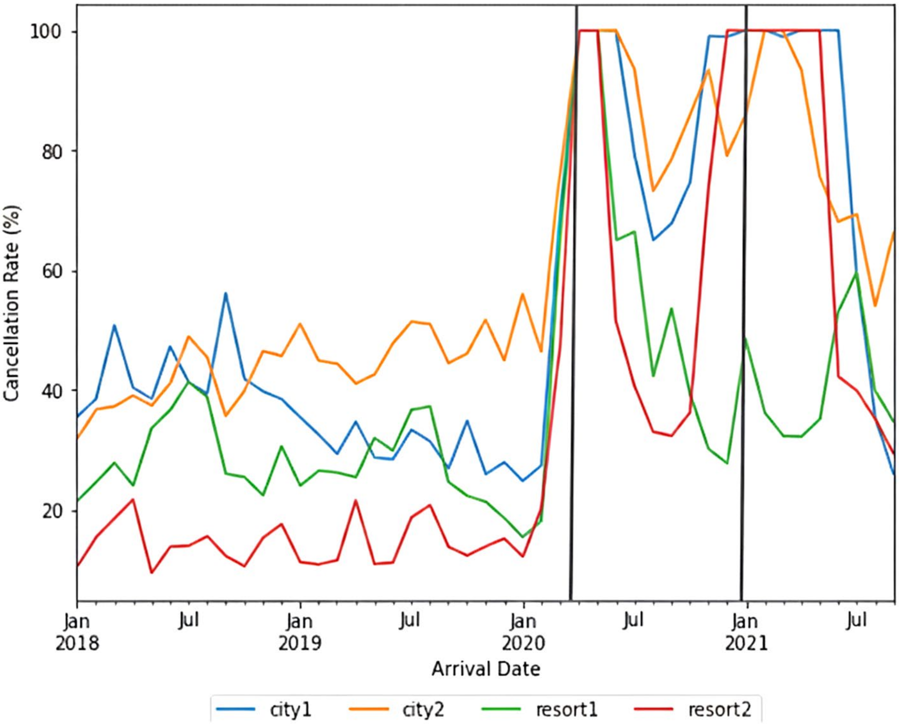

Fig. 2 Monthly cancellation rate (%) per hotel

in January, February, March, May, and June of 2021. It is important to note that in November and December 2020 and March 2021, the hotel had more than 98% cancellations. Hotel C2 had all bookings canceled for April, May, June 2020, February, and March 2021. Hotel R1 had all bookings canceled for April and May 2020. Lastly, hotel R2 has all bookings canceled for April, May, and December 2020. Then, there will be January, February, March, April, and May 2021.

Combining this information with the COVID-19 pandemic evolution regarding new infection cases and policies implemented in Portugal, one can see that following the first cases in March 2020, a State of Emergency was declared (Diário da República 2020). Public health rules dictated the closing of commercial buildings such as restaurants and bars (restaurants could operate only for take-away), the enforcement of remote working, but, most notably in terms of tourism, ordered a curfew. Overall, the State of Emergency was renewed until May 2, 2020. The State of Calamity started with measures to reopen different businesses and return to normality (Parlamento Português 2021). However, in November, a new escalation of cases made the State of Emergency mandatory again. New cases reached an all-time high in January 2021. One can relate the first significant closing of hotels to the first State of Emergency, from April to around June 2020. The second wave of cancellations in mid-2021 coincides with the start of the Delta variant in Portugal in April 2021 and a spike in new cases. Also, in June 2021, the United Kingdom (UK) added Portugal to the "red list", making quarantine mandatory for people returning from Portugal. At the same time, Portugal decided that UK citizens who were not fully vaccinated needed to quarantine for 14 days.

3.2 Data Preparation

Significant transformations of the datasets were performed using data engineering transformations that required several iterations to train models, results' evaluations, and build new transformations, reducing variable correlation and leakage and improving the final model performance.

First, features that had no contribution to the predictive power of the models or features that were identifiers were dropped, namely KardexId, FolioNumber, Meal, ReservedRoomType, AssignedRoomType, DaysInWaitingList, and RequiredCarParkingSpaces.

Secondly, categorical features were dropped due to high cardinality and a high percentage of missing values. By far, Company was the feature with the highest percentage of missing values. With 97.31% missing values for hotel C1, 95.59% for R1, and 81.01% for R2. For hotel C2, only 1.4% were missing. However, it was later found that a very high correlation existed between this feature and Agent, and thus, the decision to drop this feature was still applied.

For categorical features with high cardinality but few missing values, such as Agent, a threshold was applied to retain the more representative ones. For this feature, in Hotel C1, five unique agents were selected out of 475 unique agents. These five agents represented 60.29% of the total bookings. In Hotel C2, the top five were selected, contributing to 36.8% of the total bookings. For R1, the top five unique agents represented 51.11% of the total bookings. Lastly, R2's top five unique agents accounted for 77.16% of the total bookings. The values not belonging to the selected threshold were grouped in a new category named "Other" for all the hotels. Later, after encoding, this column was dropped.

For the DistributionChannel, the same logic was applied. For Hotel C1, there were four unique distribution channels, out of which the top two were responsible for 95.32% of the total bookings. For hotel C2, there was a record of only one unique distribution channel; thus, this column and feature were dropped. Hotel R1 had four unique distribution channels, with the top two accounting for 93.29% of the total. Lastly, for hotel R2, the top three accounted for 91.21% of the total bookings from six unique distribution channels. As in the previous case, the values not belonging to the selected threshold were grouped in a new category named "Other".

The feature MarketSegment is highly influenced by human judgment, attributing its value to a lack of consistency across hotels, and thus, the decision to drop this feature was made.

CustomerType had some highly correlated fields and was hard to interpret, for example, "Transient" and "Transient-Party". As such, the decision to create a new feature was made. The final feature, IsGroup, is a binary variable with 1 indicating a group of customers, based on the condition that there is a group when the CustomerType is "Transient-Party" or "Group".

The Country feature was also removed from the modeling datasets because it introduced leakage. The leakage was due to the default filling of Portugal as the country of origin in the bookings, information that was only confirmed and corrected at the check-in, as previous research demonstrated (António et al. 2019c). DepositType varied from hotel to hotel and was also removed.

To capture seasonality, a phenomenon of recognized importance in the tourism industry (Song and Lin 2010), ArrivalDate features were kept in the form of ArrivalDateMonth, ArrivalDateWeekNumber, and ArrivalDateDayOfMonth. The features ArrivalDate and ArrivalDateYear were removed.

The ReservationStatusDate is responsible for leakage in the dataset since all "Non-Cancelled" bookings have this field equal to the check-in date, and "Cancelled" bookings have this field equal to the date of cancellation. For the reason mentioned above, this variable was also removed.

Adults, Children, and Babies were re-engineered into a binary field "IsFamily" that outputs "1" if there are any children or babies. StaysInWeekendNights and StaysInWeekNights were re-engineered to TotalNight, recording the total number of nights, while StaysInWeekendNights was used to create IsWeekend as a binary field.

Following previous work from (António et al. 2019c), ADRThirdQuartileDeviation was differentiated from Average Daily Rate (ADR) by capturing the ADR distribution and amplitude. The reason is that bookings with a high price, compared with similar bookings for the same period, room type, and distribution channel, tend to have a higher cancellation rate. However, the ADR does not provide information about its position to similar bookings. Several modeling iterations were necessary to capture this positioning of a booking price against similar bookings to uncover an engineered feature that would incorporate price on a normalized scale for any period of the year (António et al. 2019c). The following formula calculates ADRThirdQuartileDeviation:

$$\text{ADR Third Quartile Deviation} = \frac{\text{ADR}}{\text{ADR of the third quartile of Distribution Channel}}\text{, per room type, week and year}$$

Lastly, category encoding was performed for the features ArrivalDateMonth and the selected Agents.

3.3 Modeling

3.3.1 Study one

The first study applied a classical approach of splitting the data between a defined training, validation, and testing period to analyze how well a model trained with pre-pandemic data performed in a pandemic scenario.

3.3.1.1 Data splitting and data construction Because it was a data-rich situation and following previous work in the same domain, it was employed the approach recommended by (Hastie et al. 2001) of splitting the datasets into three parts: a training set for fitting the model, a validation set for assessing the prediction error, and a test set (holdout) for assessing the generalization error (Hastie et al. 2001). The training set was generated with 67% of the pre-pandemic data before or equal to the February 29, 2020, threshold. The remaining 33% was used to generate the validation set. Considering that both the training and validation sets needed to capture a similar trend of cancellation rate, the method train_test_split was applied with stratification, thus ensuring that the proportion of canceled bookings was the same in all samples (scikit-learn.org 2021). The training set was used to train the model, and the validation set was used to fine-tune the best parameters. The test set with the "COVID-19" data was applied to assess the final model performance. Table 2 describes each hotel dataset's total canceled and not canceled bookings. Also, it represents the cancellation rate for each set.

3.3.1.2 Training As described before, hoteliers need to understand which variables most affect booking cancellation, making understanding ML models precious. Most high-performance Machine Learning (ML) techniques are black boxes that generate highly complex predictive equations (Kuhn and Johnson 2013). Nonetheless, the out-

Table 2 Total bookings per dataset sample (%Can: % canceled bookings)

	Training			Validation			Test		
Hotel	*#0*	*#1*	*%Can*	*#0*	*#1*	*%Can*	*#0*	*#1*	*%Can*
C1	69 775	48 488	41%	34 368	23 882	41%	6 643	14 729	69%
C2	65 763	50 597	43%	32 392	24 921	43%	17 118	74 178	81%
R1	42 016	16 086	28%	20 695	7 923	28%	10 651	10 778	50%
R2	19 549	3 245	14%	9 628	1 599	14%	7 258	5 481	43%

puts of some techniques, such as those based on decision trees, are more accessible for humans to understand (Abbott 2014; Hastie et al. 2001; Kuhn and Johnson 2013).

XGBoost was the chosen model, following previous work from (António et al.) in the same domain (António 2019a; António et al. 2017b, 2019c). Being a gradient-boosting-based algorithm, XGBoost is usually faster than other training model methods and allows the user to understand the importance of each feature and its contribution to the prediction of the outcome (Hastie et al. 2001).

Different feature engineering and tuning combinations were experimented with during this phase to achieve the best results. Different variations of ADR were tested, such as ADRThirdQuartileDeviation, ADR, ADR/mean, ADR/median combined with ArrivalDateWeek, or ArrivalDateMonthDay, or excluding both. Also, the hypothesis of including information from the BookingDate was tested, namely, the week number of the booking reservation date BookingDateWeekNumber. However, this experiment did not achieve better results. The best results were achieved for a validation set size of 33%, with a validation set size of 50% and 20% also being tested.

The tuning was performed using a Grid and Random Search to reduce overfitting. The best results are achieved after tuning and testing different parameters (see Table 3).

3.3.2 Study two

A dynamic approach of defining a sliding window (also known as a rolling window) was applied to assess whether incorporating data from the pandemic period improved the model. We adopt a sliding-window training design to accommodate temporal instability (concept drift) by prioritising recent observations; SHAP-based explainability then tracks rotations in feature salience across regimes. The sliding window technique allows the model to continuously update its understanding of booking patterns by training on the most recent data. This process ensures that the model stays relevant as conditions change, such as during the pandemic, when cancellation rates fluctuate. This phenomenon was precisely what happened with the pandemic, as clearly evidenced in Figs. 1 and 2. Ideally, a model can evolve and include these change patterns over time (Žliobaitė et al. 2016). Applying a sliding window to capture concept drift is an innovative approach. This method could be explained in simpler terms for a broader audience by stating that the model is periodically retrained

Table 3 Best models' tuning hyperparameters

Parameter	R1; C2; C1	R2	Values Tested
colsample_bytree	0.5	0.5	(0.4,0.1,1.3)
learning_rate	0.1	0.05	(0.03, 0.05, 0.1, 0.3)
gamma	5	5	(0,5,10)
reg_lambda	5	10	(0,5,15)
reg_alpha	5	5	(0,5,10)
subsample	0.8	0.7	(0.5,0.1,1)
max_depth	3	6	(2,1,10)
min_child_weight	1	1	(1,2,6)
scale_pos_weight	3, 1.3, 1.4	6.1	specific

with the newest available data, discarding older data to adapt to new trends. This approach is particularly justified during a pandemic due to the rapid and profound shifts in customer behavior and booking patterns.

3.3.2.1 Data splitting and data construction Sliding window techniques consist of three main steps: algorithm selection, optimal training window size definition, and prediction/test window size definition (Khan et al. 2019). For the same reasons as in study one, the chosen algorithm was XGBoost. Research shows that the ideal test set ratio is inversely proportional to the square root of the number of freely adjustable parameters. In the case of this work, it would be around 22% of the total data available, that is, the sum of training, validation, and test data, since there are 20 to 21 independent features (Amari et al. 1997). Eight combinations of train and test window size were performed, with three combinations having 24 months of training and 3, 6, and 9 months of the test set, individually. The other four combinations had 36 months of training with 3, 6, 9, and 12 months of the test set individually. Lastly, the remaining combination had 15 months of training and three months of the test. In the end, the combination of 24 months of training and nine months of the test was the best mix between performance and test set representativeness. Figure 3 shows a timeline for the selected period, showing the two windows. Window 1, W1, represented in blue, starts on 2018/03/11 and goes until 2020/11/25, with the training data going through 2020/02/29, the pandemic's start. Window 2, W2, represented in green, advances precisely six months in time, as seen by a shift in the horizontal. It starts on 2018/12/06 and ends on 2021/08/2022, with the training data going through 2020/11/25 and the testing data being the remaining data.

Table 4 Shows how the data is split between windows. One can observe that the training set for W2 has a higher cancellation rate than W1, which makes sense considering that W2 includes nine months of training data during the start of the COVID-19 pandemic

3.3.2.2 Training Following the tuning of the XGBoost in the previous chapter, using a validation set for the purpose, it was decided to keep the same parameters during

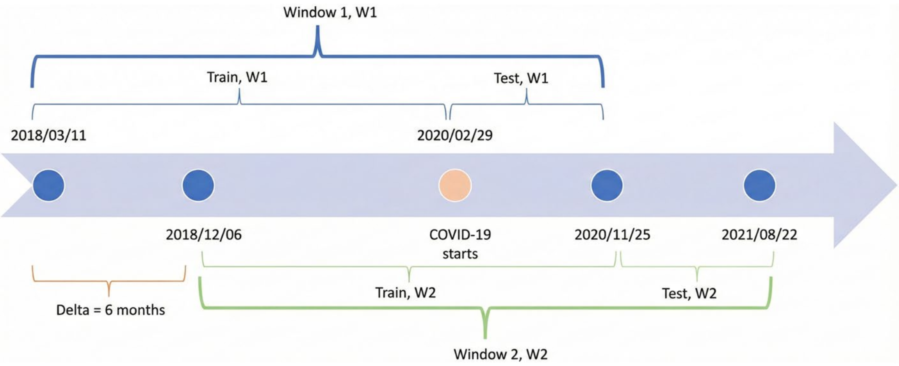

Fig. 3 Representation of the two windows

Table 4 Total bookings per dataset per sample (%Can:% canceled bookings)

		Training			Test		
Hotel	Window	*#0*	*#1*	***%Can***	*#0*	*#1*	*%Can*
C1	W1	46 700	27 471	37%	2 917	11 331	80%
C1	W2	31 221	23 852	43%	1 780	2 675	60%
C2	W1	83 375	69 469	45%	6 074	43 169	88%
C2	W2	57 313	89 786	61%	7 350	24 294	77%
R1	W1	24 937	9 934	28%	4 194	6 198	60%
R1	W2	19 278	11 663	38%	5 227	3 892	43%
R2	W1	19 915	3 502	15%	3 978	3 043	43%
R2	W2	15 330	5 075	25%	2 153	1 942	47%

this second study, with a sliding window technique. As such, there is no need for a validation set.

4 Results and discussion

This section presents and discusses the results using standard ML classification metrics - Accuracy, Precision, F1-Score (also called the F-score), and the Area Under the ROC Curve (AUC). The most popular and used evaluation metric to assess the performance of an ML algorithm is model Accuracy (Provost et al. 1998), which can be considered as the probability of success in identifying the suitable class of an observation (Maratea et al. 2014). The AUC results were considered excellent for AUC values between 0.90 and 1.00, suitable for AUC values between 0.80 and 0.90, fair for AUC values between 0.70 and 0.80, poor for values between 0.60 and 0.70, and failed for values below 0.60 (Metz 1978).

4.1 Study one

Analyzing each best-tuned version, city hotels have better results on the test set than resort hotels, with C1 having 0.77 Area Under the Curve (AUC) (fair) and C2 0.93 (excellent) compared to R1 with 0.72 and R2 with 0.70, both fair results. The test set results for the tuned versions of R1 and R2 were similar. Looking at Table 5, one can analyze the model's accuracy versus the actual incidence in the model, with R1 having 0.59 Accuracy for a 50% cancellation rate and R2 having 0.56 Accuracy for a 43% cancellation rate. C1 achieved a 0.80 Accuracy for a 69% cancellation rate, and C2 a 0.87 Accuracy for an 81% cancellation rate. The F1-Score of city hotels was 0.76 and 0.94 for hotels C1 and C2, respectively. With 0.72 and 0.70, respectively, for R1 and R2. Table 5 summarizes the performance metrics per hotel for the static model versions from Study One.

In a study comprising booking data from eight hotels in Portugal, four City hotels (Lisbon), and four resort hotels (Algarve), from 2016 to 2017, City hotels showed a 0.81 AUC vs. 0.72 in resort hotels. Another study performed with resort booking data from 2013 to 2015 achieved 0.95 AUC with boosted decision trees (BDT) (António et al. 2017a). Following this study, another study using hotel booking data

Table 5 Performance metrics per hotel and static model version (V – Vanilla (base model); T- Tuned; Acc. – Accuracy; Pre. – Precision; F1-S. – F1-Score)

		Training				Validation				Test			
		Acc.	*Pre.*	*F1-S.*	*AUC*	*Acc.*	*Pre.*	*F1-S.*	*AUC*	*Acc.*	*Pre.*	*F1-S.*	*AUC*
C1	V	0.83	0.84	0.77	0.81	0.81	0.81	0.75	0.79	0.64	0.95	0.66	0.72
	T	0.76	0.70	0.72	0.76	0.76	0.69	0.71	0.75	0.80	0.94	0.76	0.77
C2	V	0.92	0.95	0.91	0.92	0.91	0.94	0.90	0.91	0.87	0.99	0.93	0.93
	T	0.87	0.90	0.85	0.86	0.87	0.90	0.85	0.87	0.93	1.00	0.94	0.94
R1	V	0.81	0.80	0.66	0.75	0.81	0.74	0.60	0.72	0.53	0.89	0.40	0.61
	T	0.70	0.47	0.59	0.73	0.69	0.47	0.59	0.72	0.59	0.80	0.69	0.72
R2	V	0.89	0.93	0.64	0.74	0.89	0.75	0.48	0.67	0.58	0.87	0.21	0.55
	T	0.74	0.34	0.47	0.76	0.75	0.31	0.44	0.74	0.56	0.70	0.64	0.70

from 2015 to 2017 achieved 0.88 AUC on the resort and 0.92 AUC on the City hotel (António et al. 2017b). Lastly, a study using resort data from Canarias, Spain, from 2016 to 2018 achieved 0.80 AUC using random forests (RF) (Sánchez-Medina and C-Sánchez 2020. More recently, Liu et al. (2024) proposed a heterogeneous stacking-based ensemble classifier (IPHSEC) focused on profit-driven prediction, achieving high AUC values on datasets from Portugal and China, further demonstrating the potential of advanced ensemble methods in this domain. The difference between this study's results and the literature results is probably explained by the higher unpredictability introduced by the pandemic. As expected, most models built during this study show worse results than previous research (António 2019a; António et al. 2017a, b; Sánchez-Medina and C-Sánchez 2020; Liu et al. 2024). The exception is for the C2 hotel, which obtained a somewhat outstanding performance on the test set, as the hotel with the highest cancellation rate during COVID-19, 81% (see Table 6)

Before moving on to the evaluation of feature importance and its impact on the model output, it is essential to analyze the learning curves. Figure 4 shows how the Accuracy Score of the training and cross-validation evolves for an increasing number of observations in the training set. The curves are plotted with the mean scores. The shaded areas show cross-validation variability, representing a standard deviation above and below the mean for all cross-validations.

For all hotels, the validation and training scores stabilize to a value with the increasing training set size. C2 and R2 reach a cross-validation Accuracy score of near and higher than 0.80, respectively. C1 and R1 have not converged to a score higher than 0.80. Thus, adding more training instances would probably be beneficial.

Figure 5 contrasts the distributions of canceled and non-canceled bookings by LeadTime before and during COVID-19 and marks, for each hotel, the crossover threshold where the predicted probability of cancellation exceeds that of not cancelling. The thresholds compress in city hotels, from 80 to 52 days in C1 and 130 to 67 in C2, while they are essentially stable in resorts (37 to 34 in R1; 27 to 26 in R2). This pattern indicates heightened sensitivity to shorter booking horizons in urban contexts during the pandemic, whereas resort dynamics changed only marginally. This behaviour may also relate to differences in business-oriented bookings in city hotels versus leisure-focused ones in resorts, amplifying volatility in urban settings. The interpretation focuses on how these threshold shifts inform policy (free-cancellation windows, deposits, overbooking) rather than on colour regions.

4.2 Study two

This part of the discussion will focus on the results from Study Two, which employed a sliding window approach to incorporate pandemic-period data into the model training. Table 7 shows the results for the sliding window-tuned models for W1 and W2. At a first glance analysis, one can conclude that with a few minor exceptions (4 out of 32 score results), all hotels show better results on the test set than on the train set for both windows. Even though it is not frequent, this can happen when there is a high difference in the underlying distribution, which has been the case because of the start of the pandemic.

Table 6 Studies on booking cancellation prediction (Ordered by publication Year) (Avg. – Average; Min. – Minimum value; Max. – Maximum value)

Author	Data	Location	Timespan	Models	AUC on the test set		
					Avg.	Min.	Max.
António et al. 2017a	Avg. of 4 resorts	Algarve, PT	2013–2015	BDT	0.95	0.94	0.97
António et al. 2017b	Resort	Algarve, PT	2015–2017	XGB	0.88	NA	NA
	City	Lisbon, PT			0.92	NA	NA
António 2019a	Avg. of 4 Resorts	Algarve, PT	2016–2017	XGB	0.72	0.65	0.83
	Avg. of 4 City	Lisbon, PT			0.81	0.77	0.86
Sánchez-Medina & Sánchez 2020	Resort	Canarias, ES	2016–2018	RF	0.80	NA	NA
Liu et al. 2024	António et al. 2017b	Portugal, PT	2015–2017	IPHSEC	0.86	NA	NA
	CTrip	China, CH	2016	IPHSEC	0.96	NA	NA

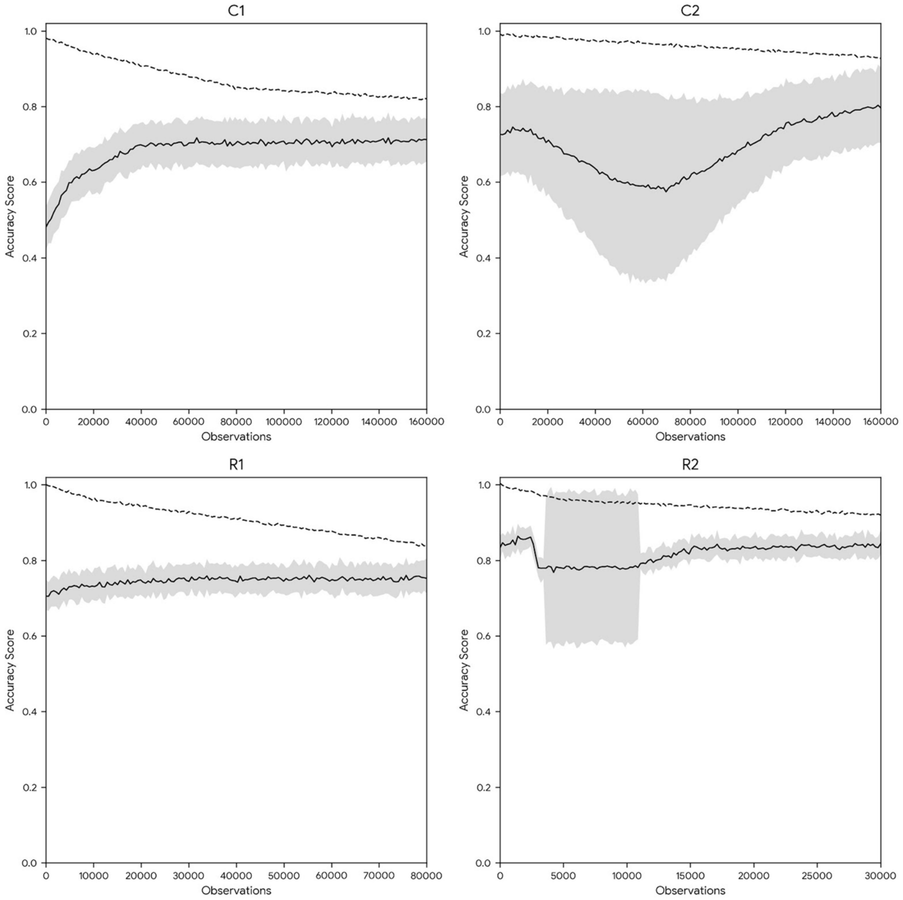

Fig. 4 Learning curve for the Vanilla (base) static models

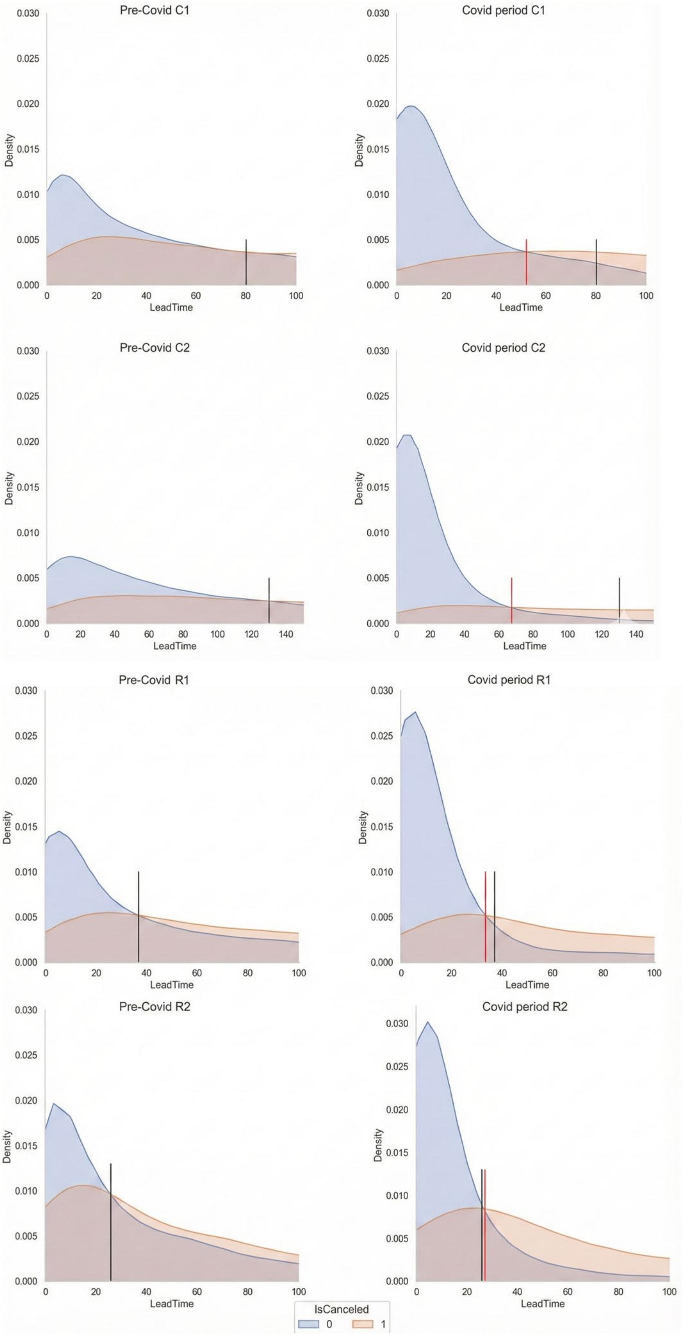

Fig. 5 - LeadTime density by hotel (pre-COVID vs. during COVID). Vertical lines denote the crossover threshold (P[cancel] > P[not cancel]); callouts display values: C1: 80 → 52 days; C2: 130 → 67; R1: 37 → 34; R2: 27 → 26. Colour shading indicates canceled vs. non-canceled distributions

Table 7 Performance metrics per hotel for the tuned sliding windows model

	Training					Test			
	Window	Acc.	Pre.	F1-S.	AUC	Acc.	Pre.	F1-S.	AUC
C1	W1	0.60	0.48	0.64	0.87	0.84	0.87	0.90	0.86
C1	W2	0.64	0.54	0.70	0.88	0.75	0.71	0.82	0.90
C2	W1	0.71	0.62	0.76	0.95	0.95	0.97	0.97	0.98
C2	W2	0.77	0.73	0.84	0.96	0.94	0.94	0.96	0.99
R1	W1	0.70	0.48	0.61	0.82	0.74	0.85	0.76	0.84
R1	W2	0.70	0.57	0.70	0.85	0.78	0.69	0.77	0.88
R2	W1	0.76	0.36	0.49	0.86	0.72	0.72	0.65	0.80
R2	W2	0.67	0.43	0.58	0.87	0.72	0.64	0.75	0.85

Second, we can conclude that when comparing W1 to W2, W2 shows better AUC on the training and test set for all hotels. In reality, all metrics display a better score on W2 than on W1, except for a few cases. This better score is probably explained by the fact that W2 includes nine months of training data since the pandemic started, thus better capturing the new patterns that arise. The exception is Accuracy in the training set for R2. For the test set, Accuracy on hotels C1 and C2, Precision on hotels C1, C2, R1, and R2, and F1-Score on hotels C1 and C2. Comparing the results from Table 7 with the %Can in Table 4, we can verify that the Accuracy on the test set is always higher than the cancellation rate on the same test set.

Comparing the results from the sliding window described in Table 7 with the results from the tuned static model, Table 5, one can verify that, for the training set, the AUC is better in both windows than in the static model. The results are particularly interesting when comparing the AUC on the test sets. Hotel C1 has a 0.86 AUC on W1 and a 0.90 AUC on W2, compared to a 0.77 AUC on the static model. Hotel C2 increased from 0.94 AUC on the static model to 0.98 on W1 and 0.99 on W2. Resorts hotels R1 and R2, which displayed the worst performance on the static model, now have relevant results on both windows. R1 increased from 0.72 AUC on the static model to 0.84 AUC on W1 and 0.88 AUC on W2. R2 increased from 0.70 AUC on the static model to 0.80 AUC on W1 and 0.85 AUC on W2. These findings are in line with previous studies that also concluded that when in the face of concept drift, using the more recent data batch for retraining the prediction model is ideal (Baier et al. 2020).

One reason for the underperformance of resort hotels compared to city hotels in Study One and their subsequent improvement in Study Two may be the higher volatility in resort bookings during the pandemic. Resort hotels, which rely heavily on seasonal and international travelers, experienced more significant uncertainty and fluctuating booking patterns, making predictions less accurate with static, pre-pandemic models. By incorporating recent pandemic data, the sliding window approach in Study Two allowed these models to adapt better to the new, volatile conditions, thus improving their performance.

Following these results, we will dive deep into W2 to better understand the performance through the learning curves and the feature behavior through the SHAP analysis. As mentioned before, the case of having a better test score than the training score suggests the model finds it easier to understand the test set. This difference in performance can happen when there is a high imbalance between sets. The learning curves shown in Fig. 6 for the W2 tuned versions show that the training score is decreasing with the growing number of training set sizes.

Figure 7 summarizes feature importance (SHAP) for the W2 models. LeadTime remains the dominant driver overall, while PreviousCancellations drops in C1 (from 4th in the static model to 13th in W2), indicating that pre-pandemic history becomes less informative under crisis. In R2, LeadTime ranks first, with TotalOfSpecialRequests, IsGroup, and rate-related measures retaining relevance. Detailed partial relationships and hotel-specific nuances are documented in Appendix B.

The SHAP dependence plots (Figs. 8, 9 and 10, and Fig. 11 in Appendix B) for key features like LeadTime, ArrivalDateMonth, ADRThirdQuartileDeviation, and TotalOfSpecialRequests for the W2 models show consistent patterns across hotels, with some variations. For instance, LeadTime generally shows a positive correlation with cancellation SHAP values. ArrivalDateMonth often shows a downward trend in SHAP values for C1, C2, and R1 as the year progresses (indicating lower cancellation risk later in the year for these models). ADRThirdQuartileDeviation also shows a generally similar pattern, although with some hotel-specific nuances.

The comparison between Study One (static model) and Study Two (sliding window model) clearly indicates that while pre-pandemic models (Study One) still offer fair to good predictive power, incorporating data from the pandemic period using a dynamic approach like the sliding window (Study Two) significantly enhances model performance, particularly for resort hotels, which were more erratically affected. This enhancement is evidenced by the increase in AUC values by up to 5% points in Study

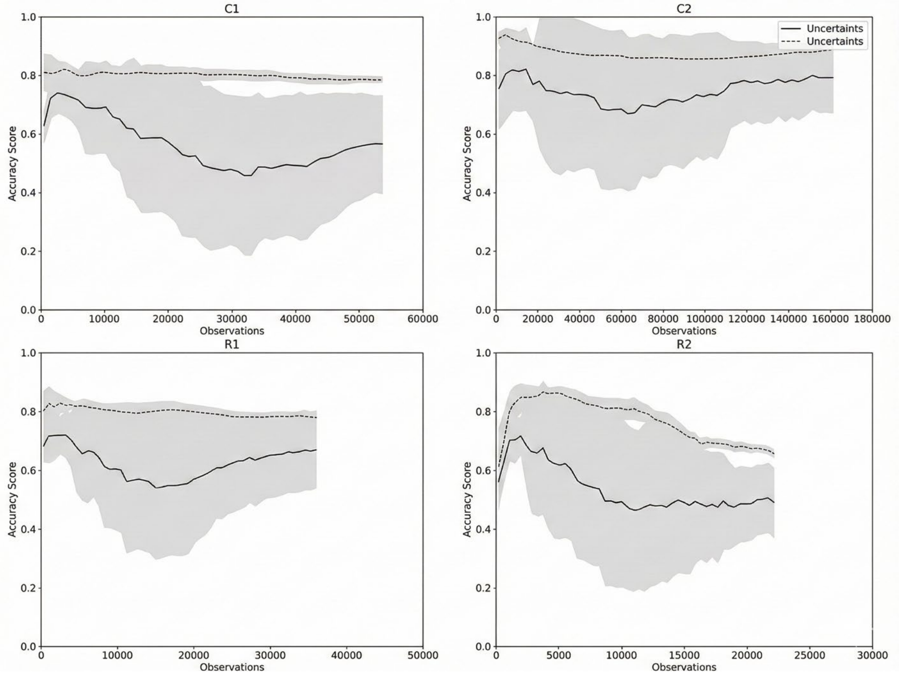

Fig. 6 Learning curve for the tuned W2 model

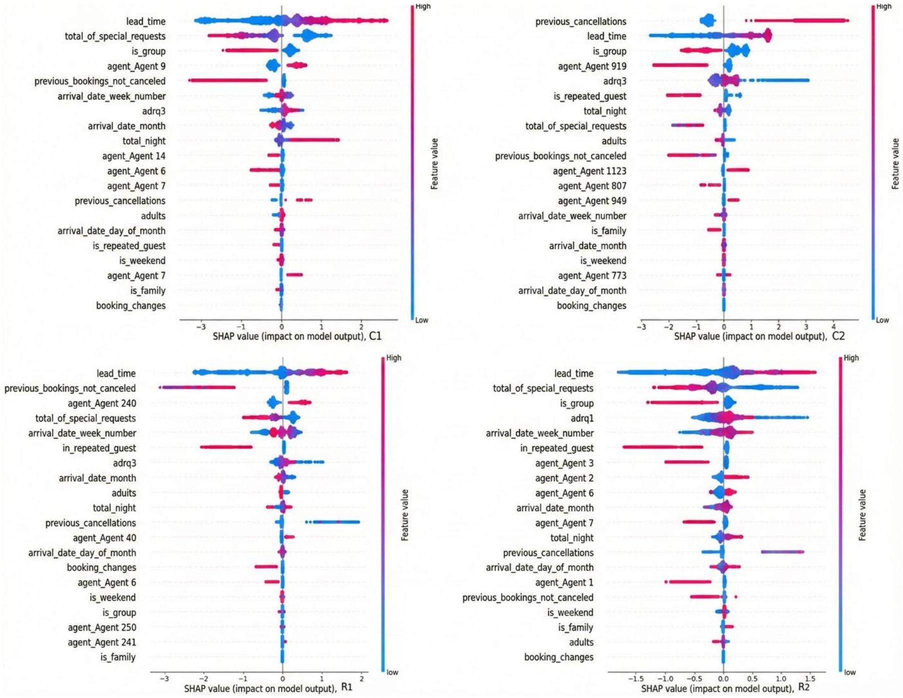

Fig. 7 SHAP summary plot for C1, C2, R1 and R2 W2 model

Two. Viewed through the lens of concept drift, the evidence indicates a regime shift in P(y|x) during COVID-19 rather than a mere change in P(x). The compression of the LeadTime crossover threshold and the rotation of feature importance (e.g., the diminished weight of pre-shock history) point to a relocated decision boundary and altered behavior. A sliding-window design serves as a drift-adaptation mechanism that improves out-of-time generalization while preserving interpretability; SHAP traces the associated shifts in feature salience. In practice, the LeadTime threshold operates as a drift diagnostic that links model behavior to policy levers (deposit timing, free-cancellation windows, overbooking) and can serve as an early-warning signal. We propose three testable propositions: (i) drift and threshold compression are stronger in urban than in resort hotels during system-wide shocks; (ii) time-varying interpretability profiles exhibit regime-consistent rotations with shock intensity; and (iii) drift-triggered adaptive retraining outperforms static schedules out of time. The shift in feature importance, especially for variables such as PreviousCancellations, underscores how customer behavior and its drivers can change during high-volatility periods, necessitating model adaptation.

5 Conclusions

This research highlights the importance of accurately estimating booking cancellations for effective demand forecasting in the hospitality industry, particularly during periods of high volatility such as the COVID-19 pandemic. The findings from two distinct studies offer valuable insights into the performance of machine learning models under such challenging conditions and how their predictive capabilities can be enhanced.

The first study demonstrates that ML classification models trained on pre-pandemic data can still achieve fair to excellent results in predicting hotel booking cancellations during volatile periods, though with slightly reduced efficacy compared to stable periods. Key predictive features like LeadTime remain crucial, though customer behavior associated with such features, like the acceptable lead time before a higher cancellation probability, altered significantly during the pandemic. The second study, employing a dynamic sliding window approach, revealed that incorporating pandemic-period data significantly improves predictive accuracy, with performance increasing up to 5% points in terms of Area Under the Curve (AUC) compared to static models. This difference underscores the critical value of dynamic model retraining with recent data under concept drift, and the importance of continuously monitoring predictive models as conditions evolve.

Beyond these results, the contributions are twofold: for theory and methods. For theory, we advance understanding of temporal instability in hospitality demand by showing how key determinants of cancellation persist yet re-parameterize under shock. LeadTime remains central while its effective threshold compresses. We make this change operational through an interpretable marker, the LeadTime crossover threshold (the point at which the predicted probability of cancellation exceeds that of not cancelling), and we document its pandemic-era compression as a compact indicator of behavioral change. For methods, we specify and assess a sliding-window training and validation protocol with explicit out-of-time evaluation, class-imbalance controls (for example, scale_pos_weight and stratified sampling) to mitigate overfitting in smaller samples, and SHAP-based interpretation to track rotations in feature importance across regimes. Together, these elements constitute a reusable blueprint for prediction under distribution shifts. From a managerial and practical perspective, this research offers actionable insights. Hotels can leverage this knowledge to identify customers who are more likely to cancel and implement preventative measures. Cancellation predictions can be integrated into revenue management systems to enhance the accuracy of demand forecasting, optimize inventory allocation, and inform pricing recommendations. Given that a higher cancellation rate directly impacts revenue, the ability to minimize cancellations in a high-volatility environment is crucial. The interpretability of models like XGBoost allows hoteliers to understand which features contribute most to cancellations, such as a long LeadTime or group bookings. This understanding facilitates the development of targeted customer management campaigns and personalized cancellation policies—for instance, offering exclusive deals to high-risk customers to retain bookings or applying less restrictive policies for low-risk profiles to encourage bookings. The observed decrease in LeadTime tolerance during the pandemic emphasizes the urgent need for hotels to constantly update safety policies and clearly communicate them to alleviate customer concerns.

Furthermore, hotel managers should consider the most recent available data to understand new booking patterns and their future implications, adapting strategies for dynamic pricing and inventory adjustments proactively rather than relying on historical data that may no longer reflect current market dynamics. Additionally, PMS developers and vendors play a pivotal role as change agents, operationalizing these ML insights into tools like integrated predictive modules for real-time forecasting. At the systems-design level, these findings translate into periodic retraining on a rolling window, drift monitoring (performance and threshold shifts), consistent feature management, and controlled model versioning/rollback within PMS pipelines. The models developed for COVID-19 enable hotels to intervene more effectively before check-in, act preemptively on potential cancellations, and generate more accurate demand forecasts. However, this research has certain limitations that open avenues for future work. One limitation was the proportion of test set data in Study One for some hotels. Future studies should aim to re-evaluate these models with larger, proportionally balanced datasets. Although the models achieved good results, they were not deployed in a live production environment. The inability to understand qualitative reasons behind cancellations or hotel-forced cancellations is another constraint; future research could incorporate features representing these aspects. The study also did not consider the customer's country of origin, which could offer further insights, particularly concerning international travel restrictions. Future work could build upon the sliding window technique to create real-time prediction models, explore explicit weighting for data recency, and analyze other hotel types and locations.

Appendix A

Feature	Type	Description
ADR	Numeric	Average Daily Rate paid by the customer or agreed to pay (if canceled) for each night of the stay
Adults	Numeric	Number of adults
Agent	Categorical	The ID of the agency (if booked through an agency)
ArrivalDate	Date	Date of arrival
AssignedRoomType	Categorical	Room type assigned to the guest
Babies	Numeric	Number of babies
BookingChanges	Numeric	Heuristic created by summing the number of booking changes (amendments) before arrival that could indicate cancellation intentions (arrival or departure dates, number of persons, type of meal, ADR, or reserved room type)
Children	Numeric	Number of children
Company	Categorical	The ID of the company/corporation (if an account was associated with it)
Country	Categorical	Country ISO identification of the primary booking holder

Feature	Type	Description
CustomerType	Categorical	Type of customer (group, contract, transient, or transient party); this last category is a heuristic built when the booking is transient but is fully or partially paid in conjunction with other bookings (e.g., small groups such as families who require more than one room)
DaysInWaitingList	Numeric	Number of days the booking was on a waiting list prior to confirmed availability and to being confirmed as a booking
DepositType	Categorical	Since hotels had different cancellation and deposit policies, a heuristic was developed to define the deposit type (non-refundable, refundable, no deposit): payment made in full before the arrival date was considered a "non-refundable" deposit, partial payment before arrival was considered a "refundable" deposit. Otherwise, it was considered "no deposit."
DistributionChannel	Categorical	Distribution channel used to make the booking
FolioNumber	Numeric	Booking ID
IsCancelled	Categorical	Outcome variable. A binary value indicating if the booking has been canceled (0: no, 1: yes)
IsRepeatedGuest	Categorical	A binary value indicating if the booking holder, at the time of booking creation, was a repeat guest at the hotel (0: no; 1: yes); was created by comparing the time of booking with the guest profile creation record
LeadTime	Numeric	Number of days before arrival that the hotel received the booking
KardexID	Numeric	The ID of the guest profile (0 if it has none)
MarketSegment	Categorical	The market segment in which the booking was classified as
Meal	Categorical	The ID of a meal the guest requested
PreviousBookingsNotCancelled	Numeric	Guest's number of previous bookings that were not canceled
PreviousCancellations	Numeric	Guest's number of previous bookings that were canceled
RequiredCarParkingSpaces	Numeric	Number of car parking spaces required by the guest
ReservedRoomType	Categorical	Room type requested by the guest
ReservationStatus	Categorical	Reservation status. Canceled or no-show should be considered Cancelled
ReservationStatusDate	Date	Date the last status was assigned to the booking. For example, canceled bookings can be used to determine how many days the booking was "alive" (based on lead time)
StaysInWeekendNights	Numeric	From the total length of stay, how many nights were on weekends (Saturday and Sunday)
StaysInWeekNights	Numeric	From the total length of stay, how many nights were on weekdays (Monday through Friday)
TotalOfSpecialRequests	Numeric	Several special requests were made (e.g., fruit basket, sea view, etc.)

Appendix B

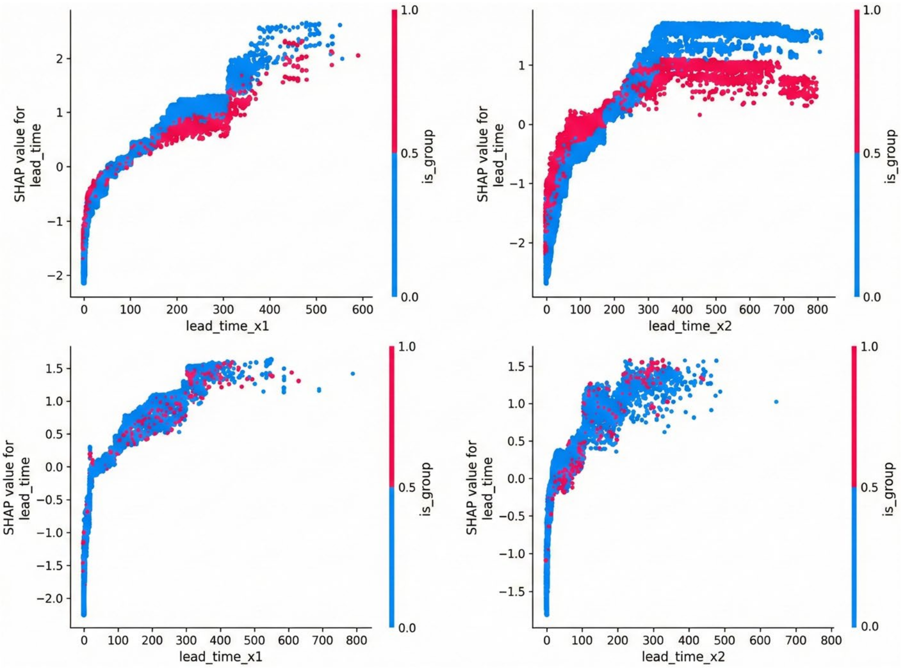

Fig. 8 SHAP dependence plot for *LeadTime* tuned W2 model

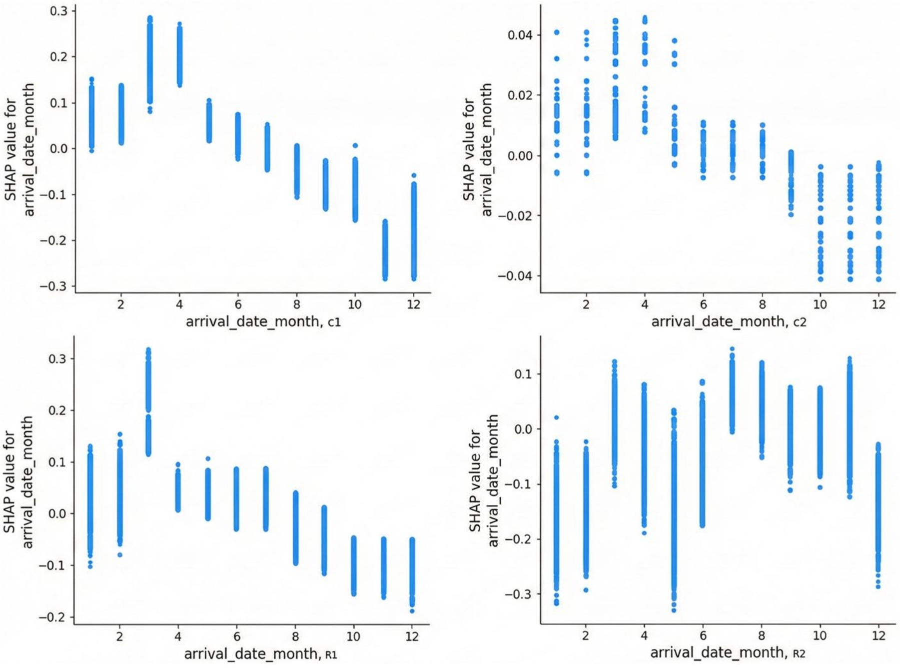

Fig. 9 SHAP dependence plot for *ArrivalDateMonth* tuned W2 model

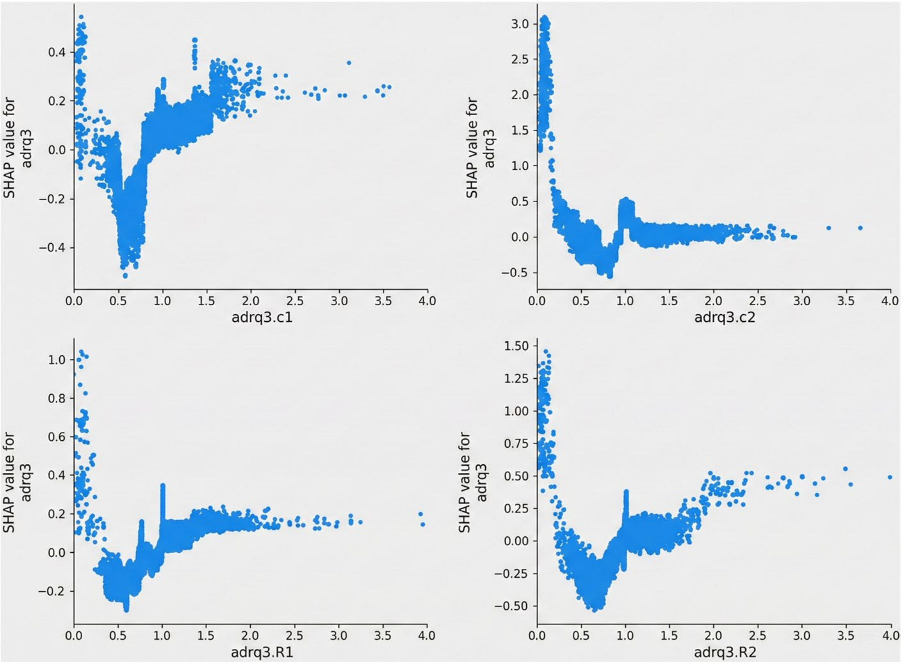

Fig. 10 SHAP dependence plot for *ADRThirdQuartileDeviation* tuned W2 model

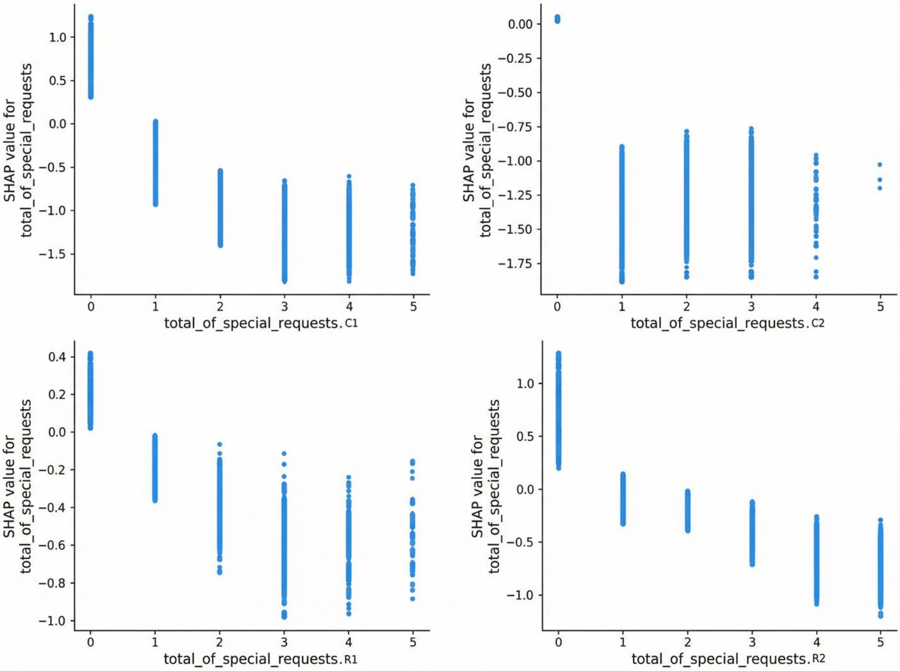

Fig. 11 SHAP dependence plot for *TotalOfSpecialRequests* tuned W2 model

Acknowledgements We thank the editors and reviewers for their suggestions on how to improve our manuscript.

Author contributions P.S. and N.A. conceptualized the research. N.A. collected the data. P.S. prepared and modeled the data. All authors analyzed the results, wrote, and revised the manuscript.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request..

Declarations

Competing interests The authors declare no competing interests.

Funding This work was supported by national funds through FCT (Fundação para a Ciê ncia e a Tecnologia), under the project - UID/04152/2025 - Centro de Investigação em Gestão de Informação (MagIC)/ NOVA IMS - https://doi.org/10.54499/UID/04152/2025 (2025-01-01/2028-12-31), UID/PRR/04152/2025 https://doi.org/10.54499/UID/PRR/04152/2025 (2025-01-01/2026-06-30), and 2024.07687.IACDC (DOI: https://doi.org/10.54499/2024.07687.IACDC).

References

Abbott D (2014) Applied Predictive Analytics: principles and techniques for the professional data analyst. Wiley. https://www.wiley.com/en-ie/Applied+Predictive+Analytics%3A+Principles+and+Techniques+for+the+Professional+Data+Analyst-p-9781118727966

Amari S, Murata N, Müller K-R, Finke M, Yang H (1997) Asymptotic statistical theory of overtraining and Cross-Validation. IEEE Trans Neural Networks 8(5):985–996. https://doi.org/10.1109/72.623200

Anguera-Torrell O, Aznar-Alarcón JP, Vives-Perez J (2021) COVID-19: hotel industry response to the pandemic evolution and to the public sector economic measures. Tourism Recreation Res 46(2):148–157. https://doi.org/10.1080/02508281.2020.1826225

António N (2019a) Big data in hotel revenue management: exploring cancellation drivers to gain insights into booking cancellation behavior. Cornell Hospitality Q 60(4):298–319. https://doi.org/10.1177/1938965519851466

António N, Rita P (2020) March 2020: 31 days that will reshape tourism. Curr Issues Tourism 24(17):1–16. https://doi.org/10.1080/13683500.2020.1863927

António N, Almeida A, Nunes L (2017a) Predicting hotel booking cancellation to decrease uncertainty and increase revenue. Tourism Manage Stud 13(2):25–39. https://doi.org/10.18089/tms.2017.13203

António N, Almeida A, Nunes L (2017b) Predicting hotel bookings cancellation with a machine learning classification model. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), 740–746. https://doi.org/10.1109/ICMLA.2017.00-11

António N, de Almeida A, Nunes L (2019b) Hotel booking demand datasets. Data Brief 41–49. https://doi.org/10.1016/j.dib.2018.11.126

António N, de Almeida A, Nunes L (2019c) An automated machine learning based decision support system to predict hotel booking cancellations. Data Sci J 18(1):32. https://doi.org/10.5334/dsj-2019-032

Baier L, Reimold J, Kühl N (2020) Handling concept drift for predictions in business process mining. 2020 IEEE 22nd Conference on Business Informatics (CBI), 76–83. https://doi.org/10.1109/CBI49978.2020.00016

Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R (2000) CRISP-DM 1.0: Step-by-step data mining guide. The CRISP-DM Consortium. https://the-modeling-agency.com/crisp-dm.pdf

Chen C-C, Zvi S, Vargas P (2011) The search for the best deal: how hotel cancellation policies affect the search and booking decisions of deal-seeking customers. Int J Hospitality Manage 30(1):129–135. https://doi.org/10.1016/j.ijhm.2010.03.010

Diário da República (2020) Decreto do Presidente da República n.o 14-A/2020. https://files.dre.pt/1s/2020/03/05503/0000200004.pdf

Ding K, Bao Y, Li L, Zhang RR, Chen Y (2025) From crisis to change: exploring the lasting influence of COVID-19 on Airbnb users through structural topic modeling. Humanit Social Sci Commun 12:1–12. https://doi.org/10.1057/s41599-025-05153-8

Falk M, Vieru M (2018) Modelling the cancellation behaviour of hotel guests. Int J Contemp Hospitality Manage 30(10):3100–3116. https://doi.org/10.1108/IJCHM-08-2017-0509

Hao F, Xiao Q, Chon K (2020) COVID-19 and china's hotel industry: Impacts, a disaster management Framework, and Post-Pandemic agenda. Int J Hospitality Manage 90:102636. https://doi.org/10.1016/j.ijhm.2020.102636

Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning. Springer. http://statweb.stanford.edu/~tibs/book/preface.ps

Hayes DK, Miller AA (2011) Revenue management for the hospitality industry. Wiley

Herrera A, Arroyo Á, Jiménez A, Herrero Á (2024) Forecasting hotel cancellations through machine learning. Expert Syst 41(9). https://doi.org/10.1111/EXSY.13608

INE (2021) Estatísticas do Turismo: 2020. Instituto Nacional de Estatística. https://www.ine.pt/xurl/pub/280866098

INE (2022) Estatísticas do Turismo: 2021. Instituto Nacional de Estatística. https://www.ine.pt/xurl/pub/22122921

Khan IA, Akber A, Xu Y (2019) Sliding window regression based short-term load forecasting of a multi-area power system. 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), 1–5. https://doi.org/10.1109/CCECE.2019.8861915

Kuhn M, Johnson K (2013) Introduction. In M. Kuhn & K. Johnson (Eds.), Applied Predictive Modeling (pp. 1–16). Springer. https://doi.org/10.1007/978-1-4614-6849-3_1

Lai IKW, Wong JWC (2020) Comparing crisis management practices in the hotel industry between initial and pandemic stages of COVID-19. Int J Contemp Hospitality Manage 32(10):3135–3156. https://doi.org/10.1108/IJCHM-04-2020-0325

Liu Z, De Bock KW, Zhang L (2024) Explainable profit-driven hotel booking cancellation prediction based on heterogeneous stacking-based ensemble classification. Eur J Oper Res. https://doi.org/10.1016/j.ejor.2024.08.026

Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. Adv Neural Inf Process Syst, 30. https://papers.nips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Luo H (2025) Understanding workforce vulnerability: the COVID-19 impact on china's hospitality and tourism labor market—evaluated by using a difference-in-difference approach. J Economic Struct 14:2485403. https://doi.org/10.1080/23322039.2025.2485403

Maratea A, Petrosino A, Manzo M (2014) Adjusted F-measure and kernel scaling for imbalanced data learning. Inf Sci 257:331–341. https://doi.org/10.1016/j.ins.2013.04.016

Matijević J, Zielinski S, Ahn Y-J (2025) Exploring the impact of COVID-19 recovery strategies in the hospitality and tourism industry. Administrative Sci 15(4):142. https://doi.org/10.3390/admsci15040142

Mehrotra R, Ruttley J, American Hotel and Lodging Association Technology Committee (2006) Revenue management. American Hotel and Lodging Association

Metz CE (1978) Basic principles of ROC analysis. Semin Nucl Med 8(4):283–298. https://doi.org/10.1016/s0001-2998(78)80014-2

Novelli M, Gussing Burgess L, Jones A, Ritchie BW (2018) No Ebola… still doomed' – The Ebola-induced tourism crisis. Annals Tourism Res 70:76–87. https://doi.org/10.1016/j.annals.2018.03.006

Parlamento Português (2021) Estado de emergência | Declarações e Relatórios. https://www.parlamento.pt/Paginas/estado-emergencia.aspx

Provost F, Fawcett T, Kohavi R (1998) The Case against Accuracy Estimation for Comparing Induction Algorithms. Proceedings of the Fifteenth International Conference on Machine Learning, 445–453

Ritchie JRB, Molinar A, C. M., Frechtling DC (2010) Impacts of the world recession and economic crisis on tourism: North America. J Travel Res 49(1):5–15. https://doi.org/10.1177/0047287509353193

Romero Morales D, Wang J (2010) Forecasting cancellation rates for services booking revenue management using data mining. Eur J Oper Res 202(2):554–562. https://doi.org/10.1016/j.ejor.2009.06.006

Sánchez-Medina AJ, C-Sánchez E (2020) Using machine learning and big data for efficient forecasting of hotel booking cancellations. Int J Hospitality Manage 89:102546. https://doi.org/10.1016/j.ijhm.2020.102546

Schwert GW (1989) Why does stock market volatility change over time? J Finance 44(5):1115–1153. https://doi.org/10.1111/j.1540-6261.1989.tb02647.x

scikit-learn.org (2021) Train_test_split from scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html

Song H, Lin S (2010) Impacts of the financial and economic crisis on tourism in Asia. J Travel Res 49(1):16–30. https://doi.org/10.1177/0047287509353190

Vong C, Rita P, António N (2021) Health-Related crises in tourism destination management: A systematic review. Sustainability 13(24):13738. https://doi.org/10.3390/su132413738

WHO (2020), March 11 WHO Director-General's opening remarks at the media briefing on COVID-19–11 March 2020. World Health Organization. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

xgboost.readthedocs (2021) XGBoost Parameters. https://xgboost.readthedocs.io/en/stable/parameter.html

Yoo M, Singh AK, Vegas L, Loewy N (2024) Predicting hotel booking cancelation with machine learning techniques. J Hospitality Tourism Technol 15(1):54–69. https://doi.org/10.1108/JHTT-07-2022-0227

Zhang Y (2019) Forecasting hotel demand using machine learning approaches. J Hospitality Tourism Res 44(2):219–244

Žliobaitė I, Pechenizkiy M, Gama J (2016) An overview of concept drift applications. In N. Japkowicz & J. Stefanowski (Eds.), Big Data Analysis: New Algorithms for a New Society (16:91–114). Springer International Publishing. https://doi.org/10.1007/978-3-319-26989-4_4

Authors and Affiliations

Pedro Silvestre[1] · Nuno Antonio[1,2] · Paulo Carrasco[2,3]

✉ Nuno Antonio
nantonio@novaims.unl.pt

Pedro Silvestre
m20200256@novaims.unl.pt

Paulo Carrasco
pcarras@ualg.pt

[1] NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, Lisbon 1070-312, Portugal

[2] CITUR – Centre for Tourism Research, Development and Innovation, Algarve, Portugal

[3] ESGHT - Escola Superior de Gestão, Hotelaria e Turismo, University of Algarve, Faro, Portugal