

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC UEH**



**ĐỀ ÁN NHÓM
KHOA HỌC DỮ LIỆU
PHÂN TÍCH TÌNH HÌNH KINH DOANH CỦA
KHÁCH SẠN VÀ GIẢI PHÁP**

Giảng viên bộ môn : Võ Thành Đức

Lớp học phần : 23D1INF50906402

Buổi học : Chiều T4

Nhóm : 2

TP. HỒ CHÍ MINH, NĂM 2023

MỤC LỤC

1. Giới thiệu đề tài	1
2. Tổng quát cơ sở dữ liệu về khách sạn ABC	1
2.1. Bộ dữ liệu	1
2.2. Tình hình kinh doanh	3
2.3. Kết luận	9
3. Bài toán dự báo	10
3.1. Mục tiêu dự báo	10
3.2. Các thuộc tính tham gia vào mô hình	10
4. Kiểm định mô hình	20
4.1. Random Forest	20
4.2. Tree	21
4.3. Logistic Regression	21
4.4. SVM	21
5. Đánh giá và lựa chọn mô hình	22
5.1. Ma trận nhầm lẫn	22
5.2. ROC	23
5.3. Test and Score	24
5.4. Kết luận	24
6. Giải pháp	25
6.1. Giữ chân khách hàng	25
6.2. Giảm thiểu thiệt hại	26
6.3. Kết luận	27
7. Phân công	28

1. Giới thiệu đề tài

Khách sạn là một hình thức kinh doanh phổ biến trên khắp thế giới nhằm đáp ứng các nhu cầu như nghỉ dưỡng, ăn uống, vui chơi giải trí và các dịch vụ khác cho những vị khách đến đây. Mô hình kinh doanh này đã tồn tại từ nhiều thế kỷ trước và để theo kịp với bước tiến hiện đại của xã hội, mô hình này đã trải qua rất nhiều biến đổi để giúp khách hàng luôn có những trải nghiệm tốt hơn khi sử dụng dịch vụ. Cụ thể, đi cùng với những tiến bộ về khoa học kỹ thuật, việc đặt phòng khách sạn đã không theo cách truyền thống thông thường mà hầu hết được đặt trực tuyến từ các kênh OTA (Online Travel Agent - dịch vụ du lịch online nhằm kết nối du khách với khách sạn và các doanh nghiệp liên quan khác), hay tại chính bản thân website của khách sạn.

Chính vì sự tiện dụng, nhanh chóng mà đặt phòng trực tuyến đã trở thành xu thế, cũng nhờ vậy nhiều khách sạn đã có một sự gia tăng đáng kể về lượt khách khi được nhiều người dễ dàng tiếp cận hơn. Tuy vậy, sự tiện dụng, dễ dàng trong hình thức đặt phòng trực tuyến cũng mang lại rất nhiều tai hại cho khách sạn, đặc biệt là tỷ lệ hủy phòng tăng cao đột biến. So với cách thức đặt phòng trực tiếp truyền thống, việc đặt phòng khó khăn hơn nên khi quyết định đặt phòng, khách hàng đã suy nghĩ và cân nhắc kỹ trước khi đặt. Trái lại, hiện tại do sự thuận tiện và đặc biệt là sự dễ dàng trong việc so sánh giữa các khách sạn, xu hướng thay đổi quyết định đặt phòng của khách hàng ngày càng tăng cao. Vì vậy, đặt phòng trực tuyến tuy đem lại một lượng lớn khách hàng nhưng cũng là lý do khiến nhiều khách hàng bỏ đi. Do vậy, nếu có những phương hướng thích hợp nhằm khắc phục tình trạng này doanh số của khách sạn sẽ được cải thiện đáng kể.

Vậy để tìm hiểu xem tình huống thực tế của ngành nghề khách sạn hiện tại ta sẽ khám phá bộ dữ liệu của khách sạn ABC về việc đặt phòng nhằm tìm ra những vấn đề mà khách sạn này đang gặp phải cũng như liệu có những phương pháp nào để khách sạn này có thể phát triển tốt hơn.

2. Tổng quát cơ sở dữ liệu về khách sạn ABC

2.1. Bộ dữ liệu

Bộ dữ liệu ghi lại thông tin các đơn đặt phòng của một khách sạn cụ thể từ tháng 7/2017 đến 12/2018 bao gồm 16 thuộc tính với 36,275 quan sát (không chứa giá trị thiếu).

```
Info
36275 instances
16 features (no missing values)
Data has no target variable.
1 meta attribute
```

Nguồn: [Kaggle](https://www.kaggle.com/datasets/abhishek1998/hotel-booking).

Mô tả thuộc tính:

Tên thuộc tính	Giải thích thuộc tính	Kiểu dữ liệu ban đầu
Booking_ID	Mã đặt phòng	Text
no_of_adults	Số người lớn	Numeric
no_of_children	Số trẻ em	Numeric
no_of_weekend_nights	Số đêm khách đặt vào cuối tuần (Thứ 7, Chủ Nhật)	Numeric
no_of_week_nights	Số đêm khách đặt trong tuần (Thứ 2 - Thứ 5)	Numeric
type_of_meal_plan	Loại bữa ăn khách hàng chọn	Categorical
required_car_parking_space	Nếu khách hàng yêu cầu có bãi đỗ xe thì là '1'; nếu không thì là '0'	Categorical
room_type_reserved	Loại phòng	Categorical
lead_time	Số ngày giữa ngày đặt phòng và ngày đến	Numeric
arrival_date	Ngày đến	Datetime
market_segment_type	Hình thức tiếp cận của khách hàng	Categorical
repeated_guest	Nếu là khách hàng cũ thì là '1'; nếu là khách hàng mới thì là '0'	Categorical
no_of_previous_cancellations	Số lần khách hàng hủy đặt phòng trước đây	Numeric

no_of_previous_bookings_not_canceled	Số lần khách hàng không hủy đặt phòng trước đây	Numeric
avg_price_per_room	Giá phòng bình quân theo ngày	Numeric
no_of_special_requests	Tổng số yêu cầu đặc biệt khác (ở tầng cao, tầm nhìn đẹp)	Numeric
booking_status	Đơn đặt phòng bị hủy hoặc không	Categorical

2.2. Tình hình kinh doanh

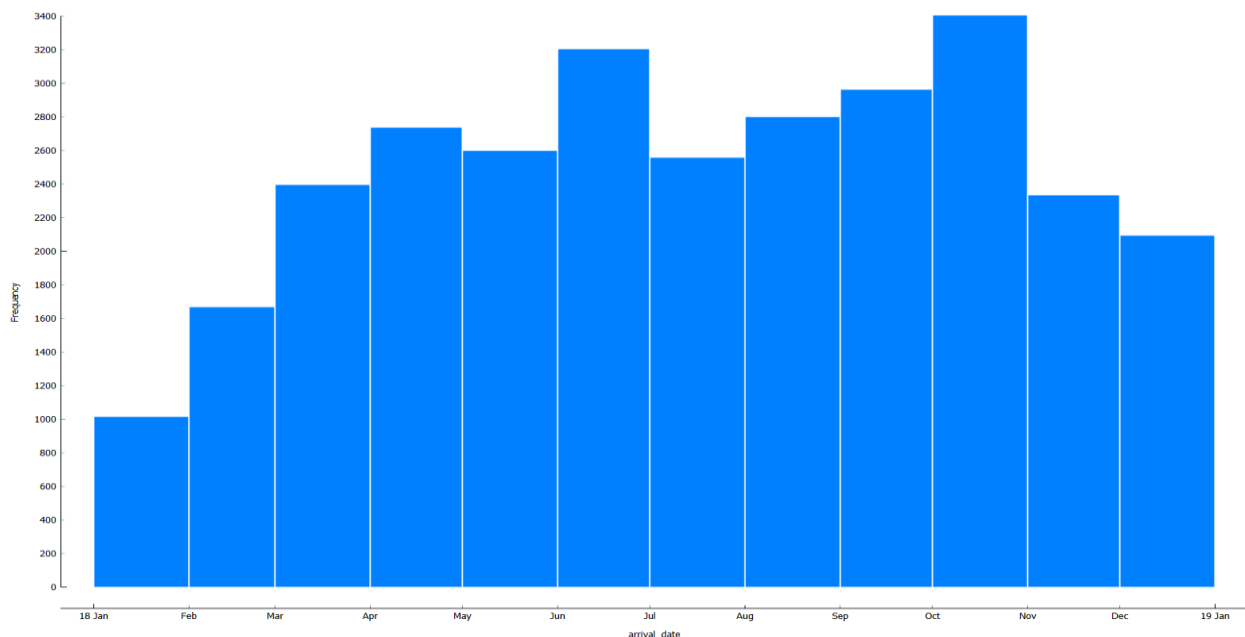
Để có thể quan sát cụ thể tình hình hoạt động trong một năm của khách sạn ta sẽ giới hạn bộ dữ liệu chỉ lấy những quan sát trong năm 2018.

Conditions

T arrival_date
is between
2018-01-01 00:00:00
and
2018-12-31 00:00:00
X

2.2.1. Hoạt động trong mùa

Như đã biết kinh doanh khách sạn là ngành nghề hoạt động theo mùa, khi vào mùa (thường trùng với mùa du lịch do khách sạn hầu như là dịch vụ không thể thiếu khi đi du lịch nghỉ dưỡng) khách sạn sẽ thu được doanh số tăng đột biến so với khoảng thời gian khác trong năm.

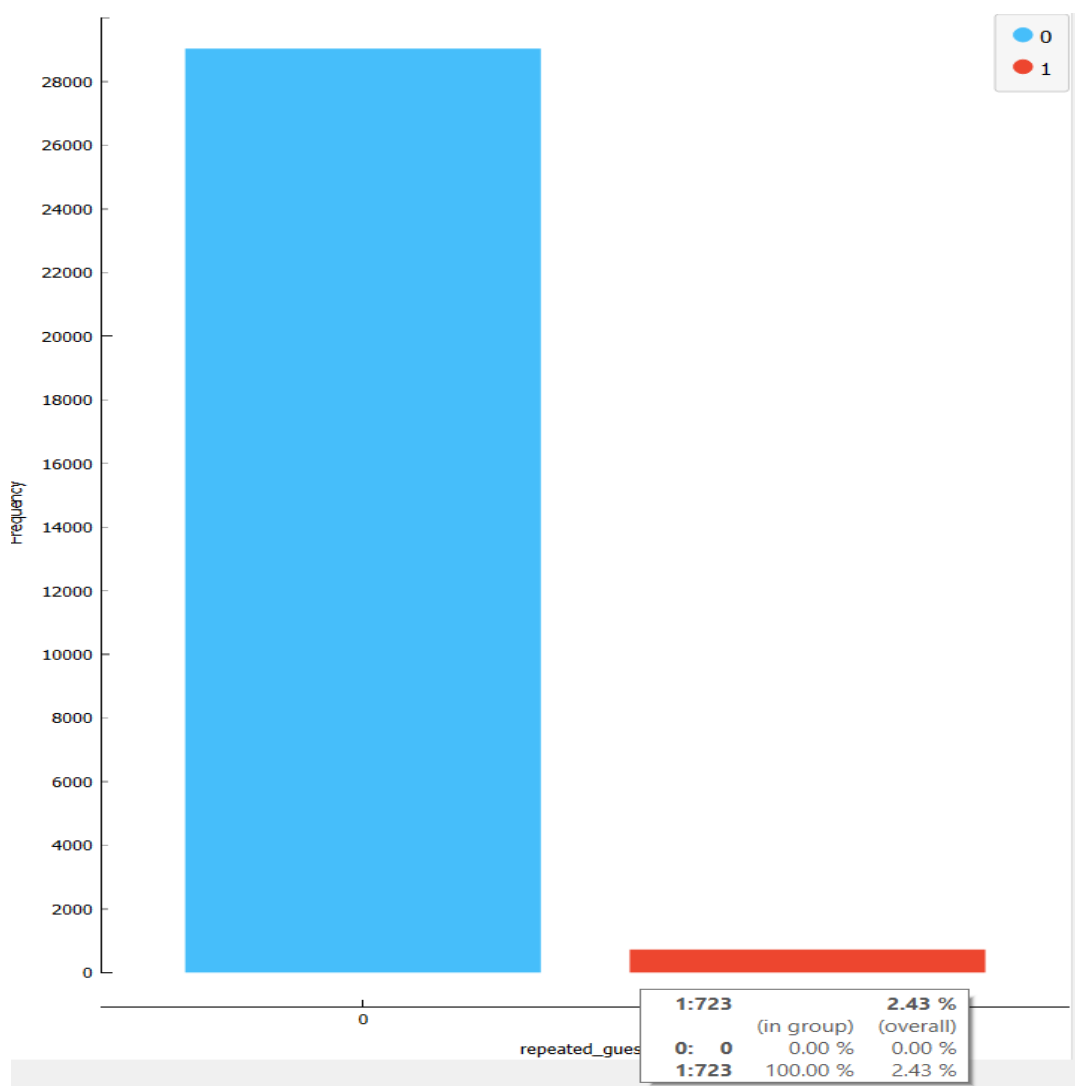


Nhìn qua hoạt động kinh doanh của khách sạn qua các tháng trong năm ta có thể thấy khách sạn có một lượng khách ổn định qua các tháng (mức chênh lệch giữa các tháng là không lớn khoảng 2-3% trên tổng số lượt đặt phòng) có thể thấy hoạt động kinh doanh không bị trì trệ. Tuy nhiên, như đã đề cập khách sạn có hoạt động kinh doanh theo mùa do vậy việc không có bất kỳ lượng khách tăng đột biến nào qua các tháng trong năm có thể nói rằng khách sạn này đang không có năng lực cạnh tranh với các đối thủ khác. Vì thế mà không thể thu được lượng lớn khách hàng nhằm tạo đột biến trong lượt khách khi đến mùa du lịch.

⇒ Các hoạt động kinh doanh của khách sạn đang không thể thu hút được khách hàng dẫn đến thất thoát một lượng khách lớn. Điều này mang lại ảnh hưởng xấu đối với doanh thu của khách sạn.

2.2.2. Lượng khách quay về

Để có thể giữ chân khách hàng quay lại khách sạn sẽ cần rất nhiều yếu tố, cũng chính vì vậy nếu tỷ lệ quay lại cao điều này đồng nghĩa chất lượng dịch vụ của khách sạn rất tốt.



Tuy vậy, đối với khách sạn này tỷ lệ quay lại của khách có thể nói là rất thấp (2.43% trong tổng số lượt khách trong năm).

- ⇒ Điều này thể hiện chất lượng dịch vụ của khách sạn cần được cải thiện nhiều hơn. Bản thân khách sạn là một loại hình kinh doanh dịch vụ do vậy việc chất lượng dịch vụ kém cũng là một trong những nguyên nhân quan trọng khiến khách sạn không có sức cạnh tranh trên thị trường này.

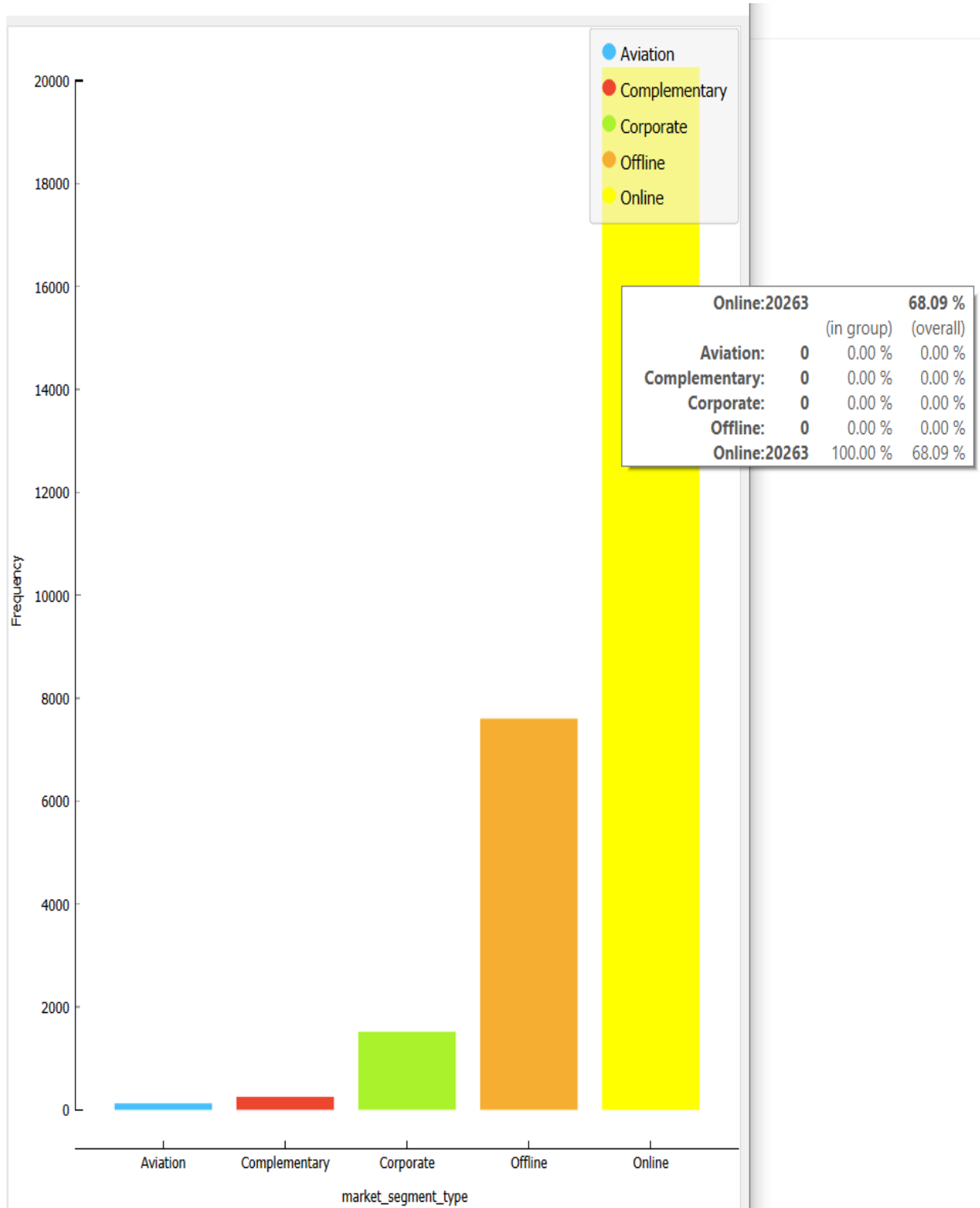
2.2.3. Tỷ lệ hủy phòng

Việc hủy phòng của khách hàng là điều mà bất kỳ nhà quản lý khách sạn nào cũng không mong muốn xảy ra. Bởi vì điều này sẽ đem lại nhiều ảnh hưởng nặng nề mang tác động tiêu cực đến doanh thu của khách sạn. Trước hết ta cần phải biết những nguồn thu lợi của khách sạn tất cả đều đến từ các dịch vụ như ăn ở (tiền phòng, buffet, ...), nghỉ dưỡng (spa, các hoạt động giải trí), tiện ích khác (đưa rước sân bay, tiện ích phòng ở, ...) và rất nhiều những dịch vụ khác. Nhưng những khoản thu này chỉ có thể nhận được với tiền đề là phải có khách, do vậy việc một khách hủy phòng sẽ mang lại rất nhiều hệ lụy cho khách sạn.

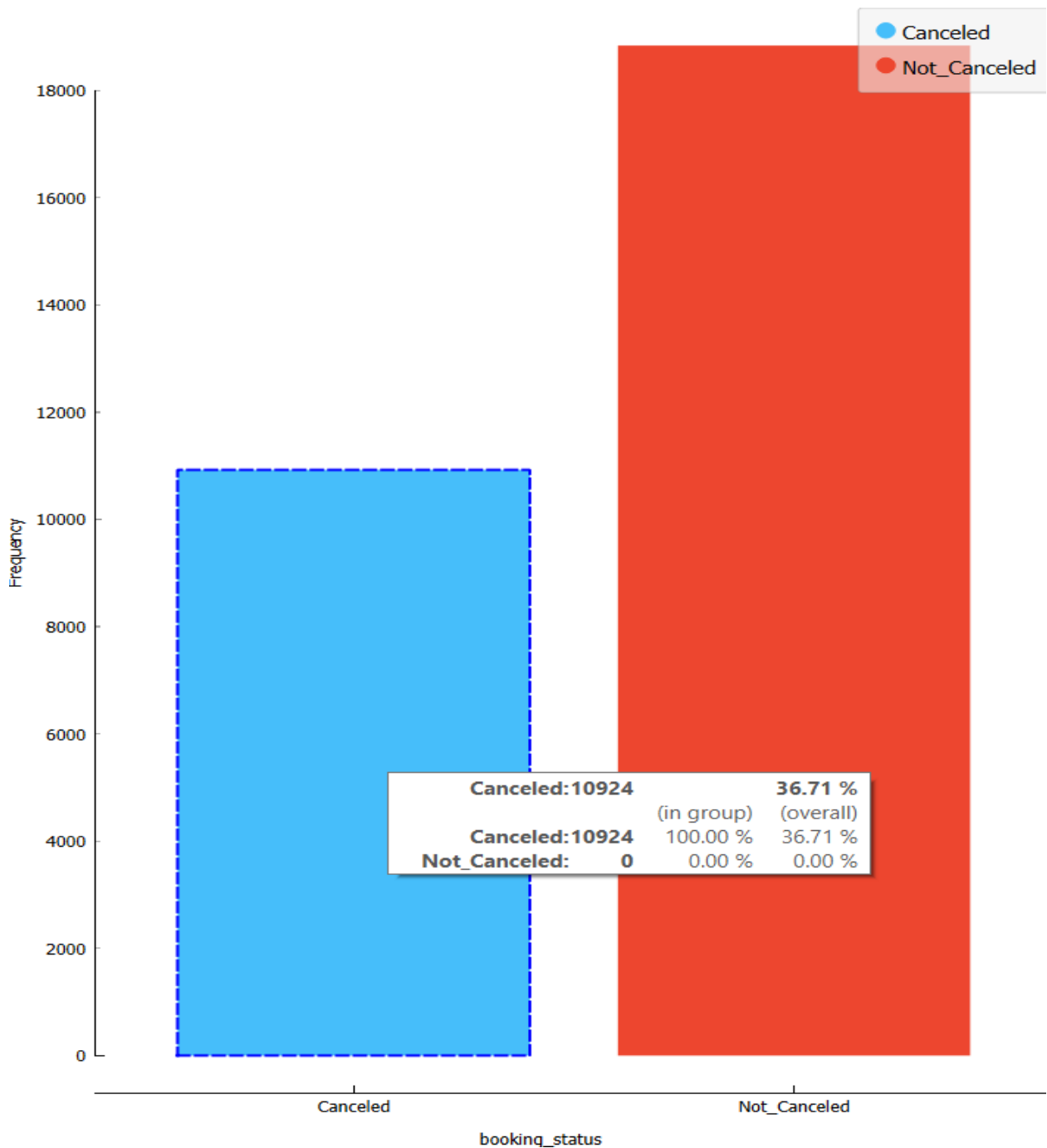
Đầu tiên khi khách hủy phòng, khách sạn sẽ mất đi những khoản thu có thể có từ vị khách này ngoài ra còn thiệt hại một khoản phí nhằm chuẩn bị khi khách đến nơi nhưng khách lại không đến. Thêm vào đó vì số phòng ở mỗi khách sạn là nguồn tài nguyên có hạn việc cho một khách đặt phòng điều này đồng nghĩa sẽ không thể nhận thêm một vị khách khác, do vậy việc hủy phòng còn làm mất đi khoản thu từ những vị khách tiềm năng đến sau.

Như đã đề cập ở phần 1 đặt phòng trực tuyến đang trở thành một xu hướng và hiện nay các khách sạn cũng đang cố gắng cải thiện để bắt kịp xu hướng này, tuy nhiên đây là con dao 2 lưỡi đối với khách sạn vừa mang về một lợi ích nhưng cũng kèm theo thiệt hại mà thể hiện rõ ràng nhất ở tỷ lệ hủy phòng của khách sạn.

Xu hướng đặt phòng trực tuyến đối với bản thân khách sạn này cũng không ngoại lệ (chiếm 68.09% trên tổng số lượt đặt phòng).



Đồng thời tỷ lệ hủy phòng của khách sạn cũng ở mức rất cao (chiếm 36.71% trên tổng số lượt đặt phòng)



⇒ Như đã nói việc hủy phòng mang đến rất nhiều hệ lụy cho khách sạn do đó với tỷ lệ hủy phòng cao như vậy sẽ là một hạn chế rất lớn đối với doanh thu khách sạn. Điều này cũng cho thấy khách sạn đang thật sự gặp khó khăn trong quá trình kinh doanh.

2.2.4. Công suất phòng

Như đã biết mọi nguồn thu của khách sạn đều cần xây dựng trên tiền đề là có khách đang lưu trú ở đây, do vậy nếu như mọi phòng trong khách sạn đều được lấp đầy khách thì khi đó khách sạn có thể sử dụng tối đa nguồn tài nguyên của mình để thu được nguồn lợi lớn nhất. Do đó ta cần một thang đo để xem khách sạn đang hoạt động như thế nào so với mức lý tưởng và đó là công suất phòng – tỷ lệ phần trăm giữa số phòng đang có khách và các phòng có thể đưa vào hoạt động của khách sạn, khi tỷ lệ này bằng 1 điều này có nghĩa các phòng của khách sạn đang hoạt động với công suất tối đa từ đó mang lại doanh thu lớn nhất cho khách sạn. Do tầm quan trọng của công suất phòng trong việc đo lường hiệu quả kinh doanh của khách sạn vì vậy việc khiến công suất phòng đạt mức tối ưu nhất là một bài toán gây trở ngại với nhiều nhà quản lý khách sạn.

Ta thấy được rằng công suất phòng sẽ phụ thuộc vào 2 biến là số phòng đang có khách (gọi là x) và số phòng tối đa của khách sạn (gọi là y). Như vậy để có thể tối ưu được công suất phòng ta cần tăng x hoặc giảm y , tuy nhiên việc giảm số phòng tối đa của khách sạn để tăng công suất phòng là bất hợp lý cho việc gia tăng doanh thu, do vậy điều duy nhất có thể làm để đạt mức công suất phòng tối ưu là gia tăng số phòng đang có khách hay nói cách khác là cố gắng để có thể kiếm nhiều khách nhất có thể. Tuy vậy việc kiếm một vị khách mới không hề dễ dàng thay vào đó nếu ta có thể tận dụng lượt khách đến từ việc hủy phòng thì đây sẽ là một kho tàng lớn để khai thác.

- ⇒ Như đã đề cập việc hủy phòng gây ra rất nhiều hệ lụy cho khách sạn do vậy ta cần có những phương án để giảm thiểu thiệt hại này đến mức tối đa. Tuy nhiên khách sạn có thể thu được nhiều lợi ích hơn nếu ta có thể thay đổi suy nghĩ của những vị khách này.
- ⇒ Vì khi họ quyết định đặt phòng điều đó có nghĩa họ đã có ấn tượng tốt với khách sạn nhưng vì một lý do nào đó họ mới hủy phòng (có thể có chuyện đột xuất cũng có thể là tìm thấy một khách sạn tốt hơn) như vậy nếu có thể tìm ra cách thức nhằm thu hút họ thì ta hoàn toàn có thể biến họ trở thành khách hàng thực sự của khách sạn. Cụ thể với tỷ lệ hủy phòng hiện tại là 36.71% với mỗi phần trăm tỷ lệ này chuyển sang tỷ lệ không hủy phòng thì công suất phòng sẽ tăng lên một mức tương ứng từ đó cải thiện doanh thu cho khách sạn.
- ⇒ Như vậy thay vì tìm cách thu hút những vị khách chưa có ấn tượng gì với khách sạn, ta nên tập trung vào những phương thức nhằm khai thác lượng khách hàng đã có sẵn ấn tượng tốt với chúng ta, đây sẽ là một nguồn tài nguyên tuyệt vời cho khách sạn từ đó giải quyết vấn đề tối ưu công suất phòng rồi dẫn đến tối ưu doanh thu của khách sạn.

2.3. Kết luận

Từ đây ta có thể thấy tình hình kinh doanh của khách sạn này đang thực sự gặp rất nhiều khó khăn do vậy ta cần một giải pháp nhằm cải thiện tình hình này. Như đã đề cập công suất phòng là một tiêu chí quan trọng nhằm đánh giá hiệu quả hoạt động kinh doanh của khách sạn do vậy cố gắng nâng cao tiêu chí này cũng là cách để cải thiện tình hình kinh doanh của khách sạn và để tối ưu được công suất phòng ta cần có những biện pháp nhằm chuyển những vị khách từ nhóm hủy phòng sang không hủy phòng.

Nhưng trước hết ta cần có cách thức để nhận diện những vị khách hủy phòng để áp dụng những biện pháp này lên họ. Tuy nhiên nếu muốn nhắm đến nhóm khách hàng đã thật sự hủy phòng rồi là không khả thi vì khi đó họ đã kiếm được một nơi tốt hơn để nghỉ dưỡng do vậy sẽ rất khó khăn để lôi kéo họ quay về. Vì lẽ đó ta cần dự đoán trước liệu vị khách nào có khả năng sẽ hủy phòng để có những phương hướng thích hợp nhằm giữ chân họ ở lại. Đồng thời ta cũng có thể hạn chế những thiệt hại do những khách hàng này gây ra nếu không thể giữ chân họ lại.

3. Bài toán dự báo
















3.1. Mục tiêu dự báo

Như đã nói để có thể cải thiện tình hình kinh doanh hiện tại của khách sạn ta cần nâng cao công suất phòng bằng cách giữ chân những khách hàng có khả năng hủy phòng và để được như vậy trước hết ta cần dự đoán xem liệu vị khách nào có khả năng hủy phòng từ đó có những quyết định hợp lý với nhóm khách hàng này. Do vậy ta sẽ xây dựng mô hình máy học dựa vào bộ dữ liệu trên để giải quyết bài toán dự đoán trên.

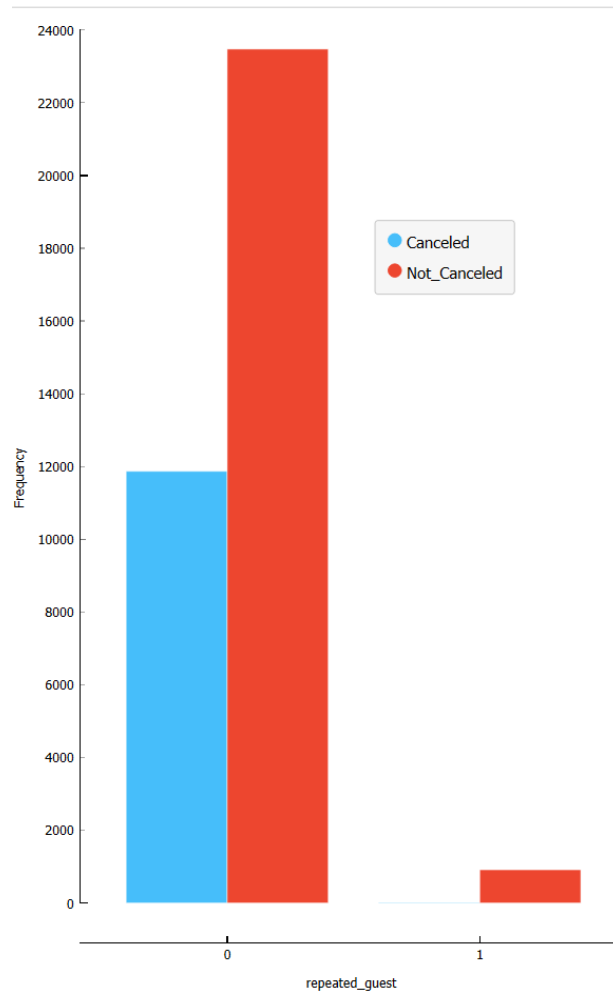
⇒ Biến mục tiêu là “booking status”.

3.2. Các thuộc tính tham gia vào mô hình

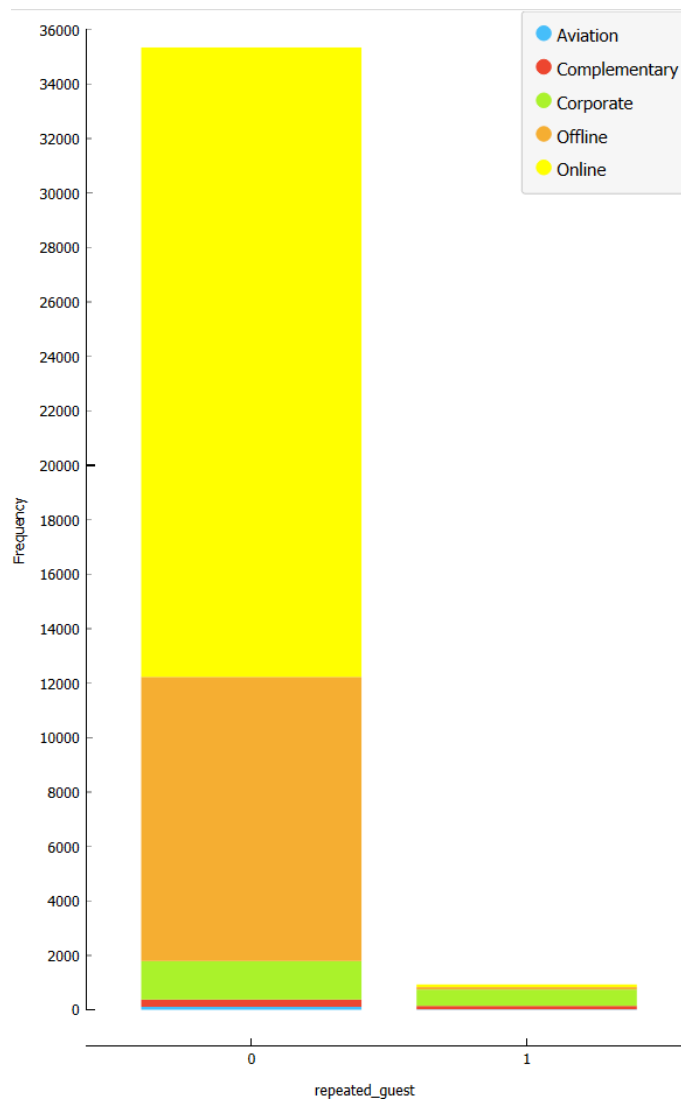
Sử dụng Rank Widget của Orange cho bộ dữ liệu này, ta chọn những thuộc tính có Gain Ratio (chỉ số đo mức tăng thông tin của biến thông tin) lớn hơn 0.01. Dựa vào đây ta có 9 thuộc tính sẽ tham gia vào mô hình dự đoán như sau:

		#	Gain ratio	Gini
1	 repeated_guest	2	0.071	0.005
2	 lead_time		0.063	0.076
3	 no_of_previous_bookings_not_canceled		0.061	0.005
4	 no_of_previous_cancellations		0.038	0.001
5	 no_of_special_requests		0.037	0.029
6	 required_car_parking_space	2	0.033	0.003
7	 arrival_date		0.022	0.025
8	 market_segment_type	5	0.016	0.010
9	 avg_price_per_room		0.012	0.014
10	 no_of_adults		0.006	0.004
11	 type_of_meal_plan	4	0.005	0.003
12	 no_of_week_nights		0.003	0.004
13	 no_of_weekend_nights		0.002	0.002
14	 no_of_children		0.002	0.001
15	 room_type_reserved	7	0.001	0.001

3.2.1. Thuộc tính `repeated_guest`

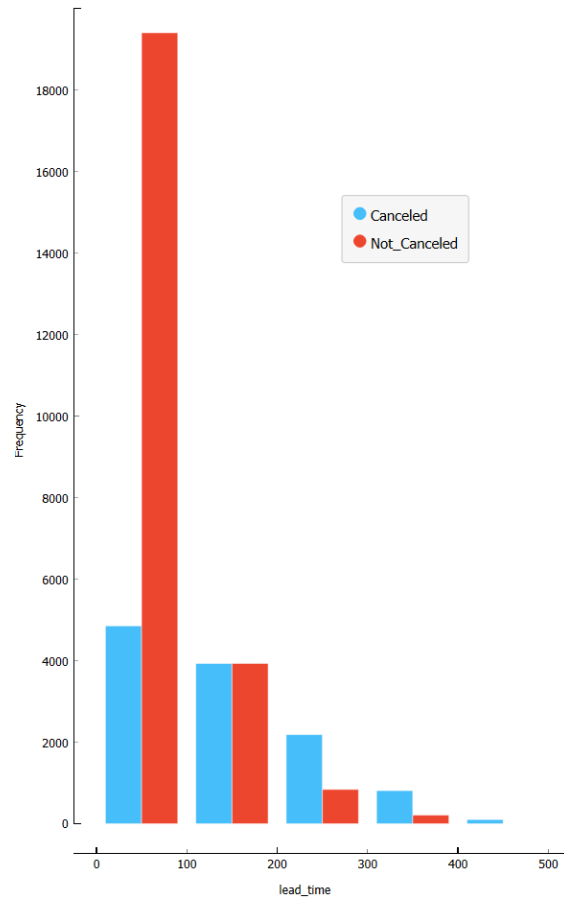


Những khách hàng đã từng lưu trú tại khách sạn trước đó có xu hướng ít hủy phòng ở lần tiếp theo hơn. Cụ thể, trong lượng khách hàng mới, tỉ lệ khách hủy phòng lên tới gần 34%, còn tỉ lệ hủy phòng của những khách hàng cũ chỉ xấp xỉ 2%.



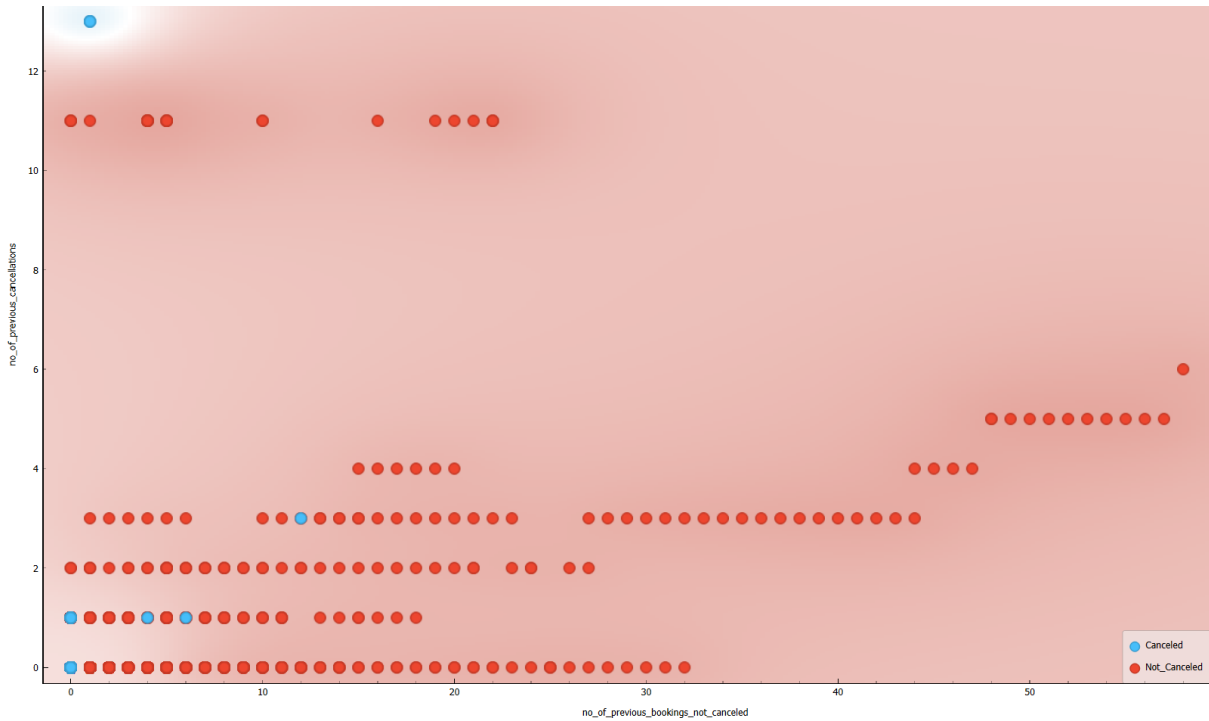
Trong số các khách hàng trở lại khách sạn, loại khách Corporate (nhóm khách thuộc công ty, doanh nghiệp) chiếm đa số với tỉ lệ là 64.73%. Nhóm khách này thường trở lại những điểm đến giống nhau mỗi năm vì mục đích công việc. Mục tiêu chọn khách sạn của nhóm khách hàng này thường là gần vị trí trung tâm hoặc nơi hội họp, kết nối wifi mạnh, nhiều điểm sạc thiết bị điện tử, chính sách ưu đãi tốt đối với khách đoàn... Và có lẽ khách sạn đã đáp ứng được những yêu cầu đó nên nhóm khách này chiếm đa số trong tỉ lệ trở lại.

3.2.2. Thuộc tính lead_time



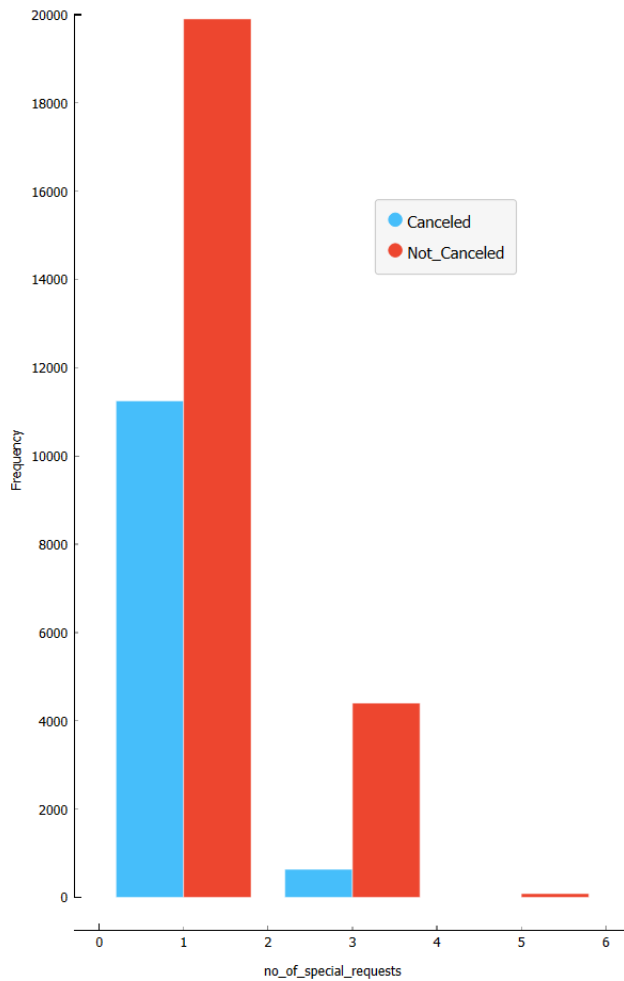
Số ngày chênh lệch giữa ngày đặt và ngày nhận phòng càng lớn thì vị khách đó càng có nguy cơ hủy đặt phòng. Cụ thể, với số ngày đặt trước nhỏ hơn 100, tỉ lệ hủy đặt phòng chỉ chiếm 20%; khi số ngày đặt trước lớn hơn 100, tỉ lệ này tăng lên hơn 50%. Mối tương quan này được thể hiện rõ khi số ngày đặt trước lớn hơn 400, tỉ lệ khách hủy phòng lên đến 100%. Việc số ngày đặt trước càng lớn càng làm tỉ lệ hủy phòng tăng cao có thể vì nhiều lí do như booking ảo, khách hàng thay đổi lịch trình, khách hàng đã chọn khách sạn khác,... Để giảm tỉ lệ hủy phòng, khách sạn nên tạo ràng buộc trong khâu đặt phòng, cụ thể là ngày đặt không cách ngày nhận phòng quá 100 ngày. Đối với những đơn có số ngày đặt phòng trước quá nhiều so với ngày nhận phòng, khách sạn cần giữ liên lạc và lên lịch nhắc khách hàng, bên cạnh đó cũng nên có những chính sách ưu đãi khi khách hàng đến nhận phòng đúng lịch, những giải pháp này sẽ được đề cập cụ thể ở chương 6.

3.2.3. Thuộc tính no_of_previous_bookings_not_canceled và thuộc tính no_of_previous_cancellations



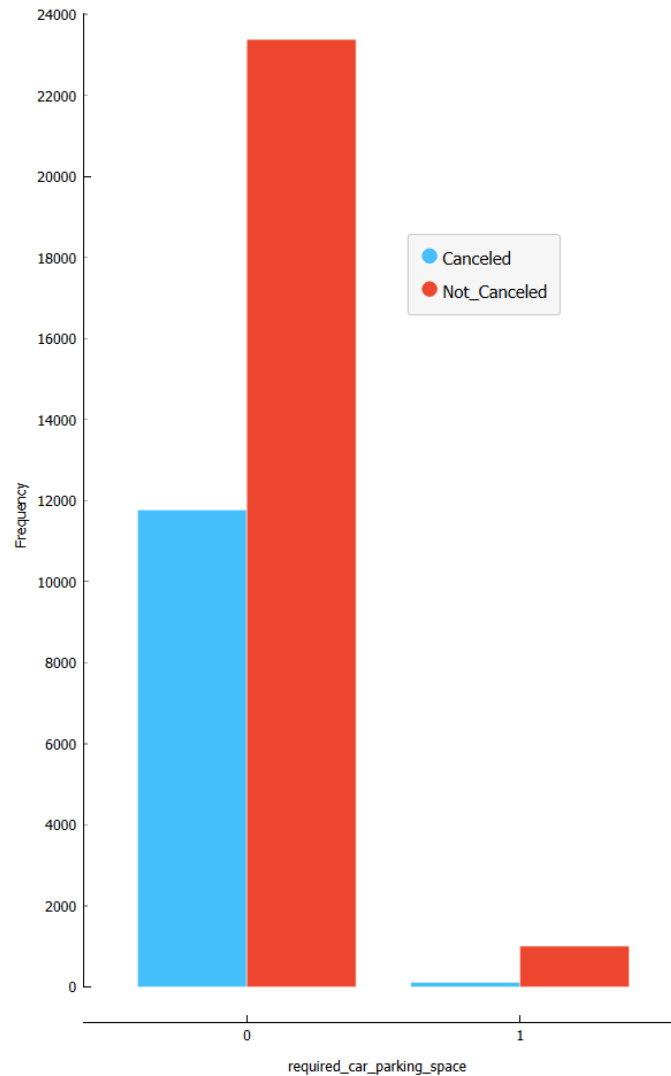
Những khách hàng có số lần hủy đặt phòng và số lần không hủy đặt phòng đều bằng 0 có thể là vì họ chưa từng sử dụng dịch vụ đặt phòng khách sạn trước đó. Những vị khách hàng này sẽ có xu hướng hủy đặt phòng ở lần đầu tiên. Những khách hàng có số lần không hủy đặt phòng càng nhiều (>15 lần) thì càng ít nguy cơ hủy đặt phòng ở hiện tại. Tuy nhiên, không phải vị khách nào đã từng hủy đặt phòng nhiều thì cũng sẽ hủy đặt phòng ở hiện tại, chỉ có 1 trường hợp là số lần hủy đặt phòng lên đến 13 lần là thật sự hủy đặt phòng.

3.2.4. Thuộc tính no_of_special_requests



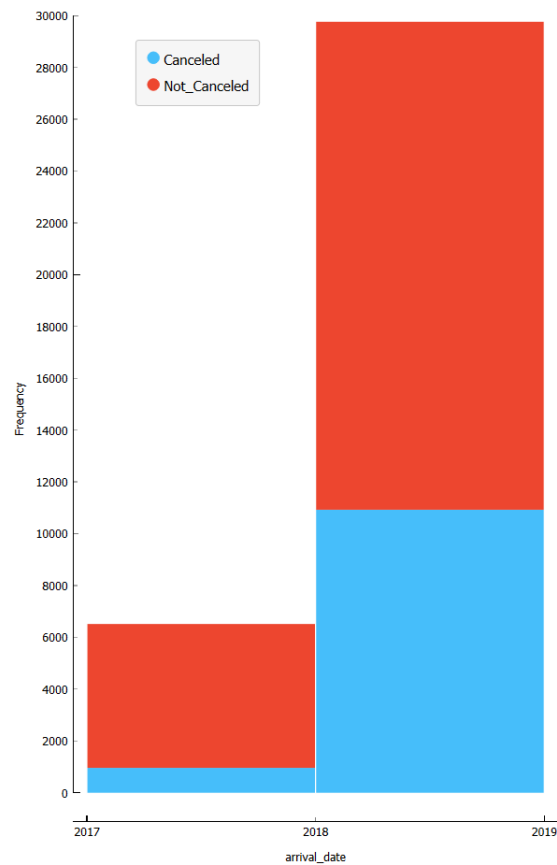
Những đơn đặt phòng càng có nhiều yêu cầu đặc biệt thì tỉ lệ hủy phòng càng thấp. Với những đơn đặt phòng không có yêu cầu đặc biệt hoặc chỉ có 1 yêu cầu thì tỉ lệ hủy đặt phòng là xấp xỉ 36%, với những đơn có nhiều hơn 4 yêu cầu thì tỉ lệ hủy đặt phòng là 0%.

3.2.5. Thuộc tính `required_car_parking_space`



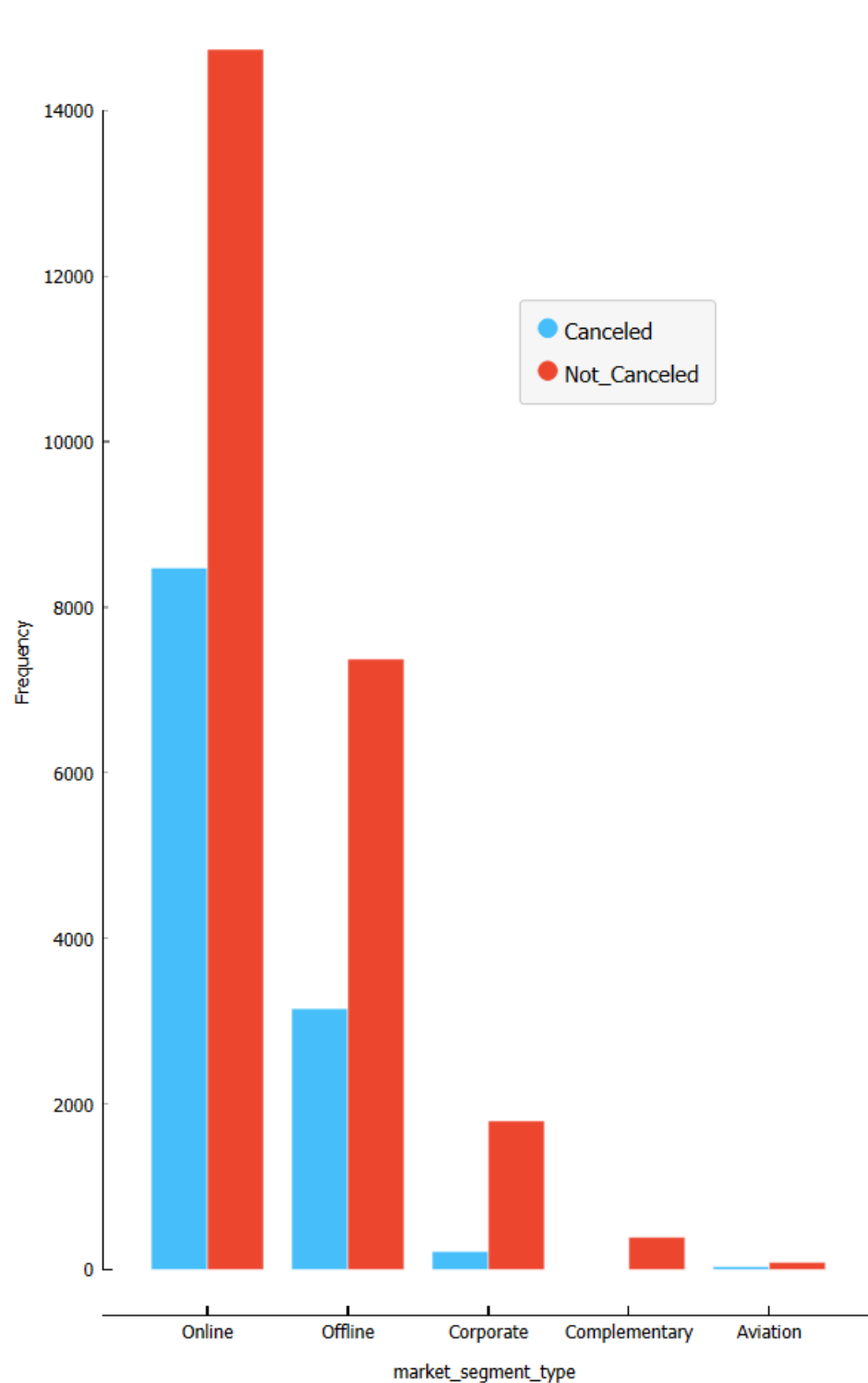
Thuộc tính này cũng có tác động đến biến mục tiêu tương tự như thuộc tính `no_of_special_requests`, những khách hàng có yêu cầu bãi đậu xe hơi sẽ ít hủy đặt phòng hơn. Trong khi những vị khách không yêu cầu bãi đậu xe có tỉ lệ hủy đặt phòng lên đến 33.5% thì tỉ lệ hủy đặt phòng của những vị khách yêu cầu có bãi đậu xe chỉ là 10%.

3.2.6. Thuộc tính arrival_date



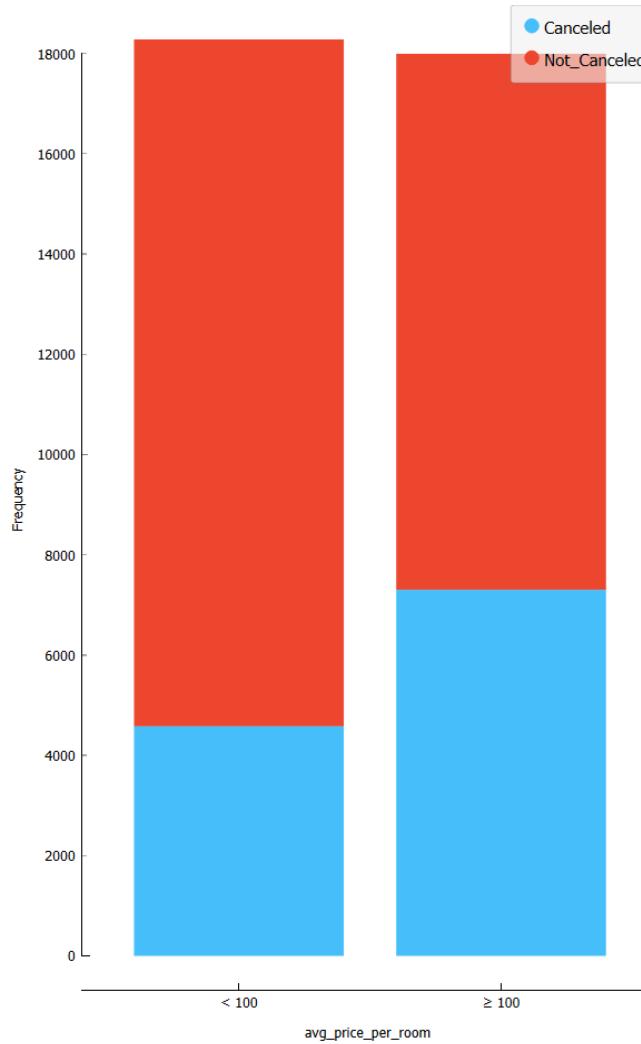
Trong năm 2018, số đơn đặt phòng khách sạn tăng 4.5 lần so với năm 2017. Tuy nhiên, tỉ lệ hủy đặt phòng trong năm 2018 lại rất cao so với 2017. Cụ thể, tỉ lệ hủy phòng năm 2017 là 14.75% nhưng tỉ lệ hủy phòng năm 2018 lại là 36.71%. Chính vì tỉ lệ hủy phòng tăng cao như vậy, khách sạn cần nhanh chóng áp dụng nhiều biện pháp để khắc phục tình hình trong những năm tiếp theo.

3.2.7. Thuộc tính market_segment_type



Như đã đề cập, hình thức đặt phòng online vô cùng tiện lợi và nhanh chóng, vì thế tỉ lệ hủy đặt phòng dưới hình thức này cũng rất cao (36.5%).

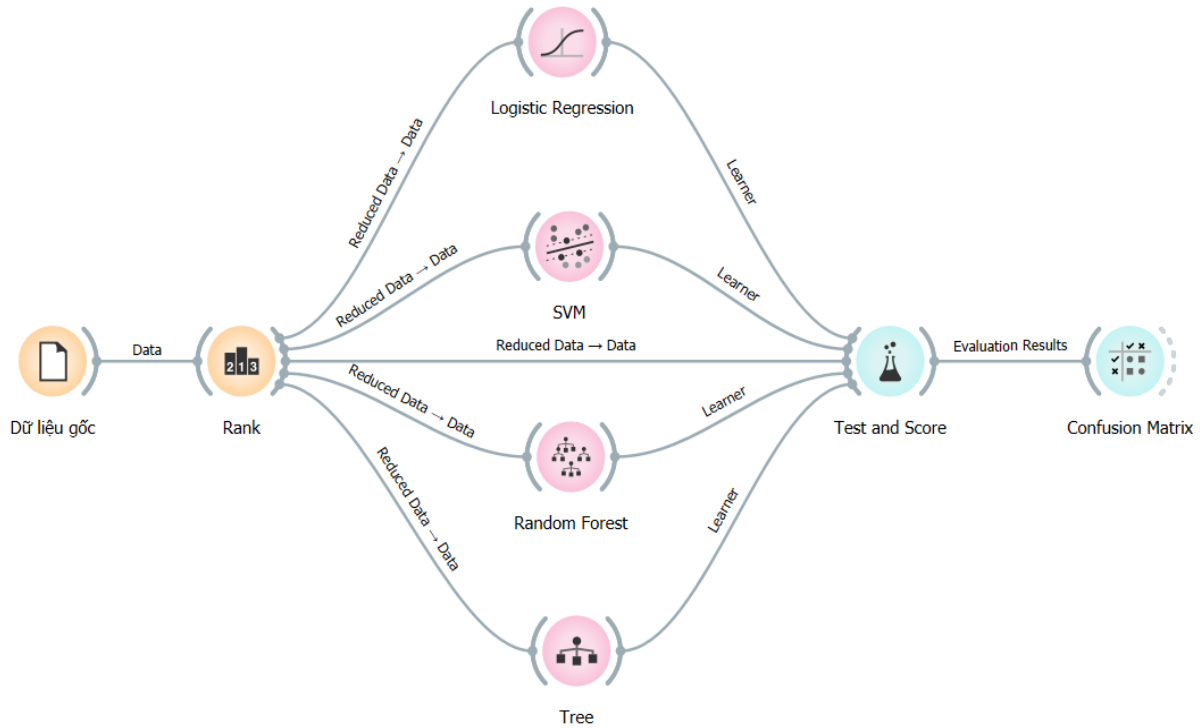
3.2.8. Thuộc tính avg_price_per_room



Giá phòng trung bình càng cao thì càng có nguy cơ khách hủy đặt phòng. Khi giá phòng dưới 100€, tỉ lệ hủy đặt phòng là 25.06%; khi giá phòng từ 100€ trở lên, tỉ lệ hủy đặt phòng là 41%.

4. Kiểm định mô hình

Sau khi đã lựa chọn được biến mục tiêu và các thuộc tính để đưa vào mô hình dự đoán, sử dụng 4 phương pháp dự đoán là Random Forest, Tree, Logistic Regression và SVM.



4.1. Random Forest

		Predicted		
		Canceled	Not_Canceled	Σ
Actual	Canceled	9591	2294	11885
	Not_Canceled	1488	22902	24390
Σ		11079	25196	36275

Một trận cho thấy số trường hợp bị nhầm lẫn của phương pháp này. Cụ thể, số quan sát có giá trị thực tế là “Canceled” nhưng lại đoán thành “Not_Canceled” là 2,294 (chiếm 6.3% tổng quan sát); số quan sát có giá trị thực tế là “Not_Canceled” nhưng lại đoán thành “Canceled” là 1,488 (chiếm 1.68% tổng quan sát).

4.2. Tree

		Predicted		Σ
		Canceled	Not_Canceled	
Actual	Canceled	9481	2404	11885
	Not_Canceled	2570	21820	24390
Σ		12051	24224	36275

Mã trận cho thấy số trường hợp bị nhầm lẫn của phương pháp này. Cụ thể, số quan sát có giá trị thực tế là “Canceled” nhưng lại đoán thành “Not_Canceled” là 2,404 (chiếm 6.63% tổng quan sát); số quan sát có giá trị thực tế là “Not_Canceled” nhưng lại đoán thành “Canceled” là 2,570 (chiếm 7.08% tổng quan sát).

4.3. Logistic Regression

		Predicted		Σ
		Canceled	Not_Canceled	
Actual	Canceled	7095	4790	11885
	Not_Canceled	2602	21788	24390
Σ		9697	26578	36275

Mã trận cho thấy số trường hợp bị nhầm lẫn của phương pháp này. Cụ thể, số quan sát có giá trị thực tế là “Canceled” nhưng lại đoán thành “Not_Canceled” là 4,790 (chiếm 13.2% tổng quan sát); số quan sát có giá trị thực tế là “Not_Canceled” nhưng lại đoán thành “Canceled” là 2,602 (chiếm 7.17% tổng quan sát).

4.4. SVM

		Predicted		Σ
		Canceled	Not_Canceled	
Actual	Canceled	2853	9032	11885
	Not_Canceled	2241	22149	24390
Σ		5094	31181	36275

Mã trận cho thấy số trường hợp bị nhầm lẫn của phương pháp này. Cụ thể, số quan sát có giá trị thực tế là “Canceled” nhưng lại đoán thành “Not_Canceled” là 9,032 (chiếm 24.9%

tổng quan sát); số quan sát có giá trị thực tế là “Not_Canceled” nhưng lại đoán thành “Canceled” là 2,241 (chiếm 6.18% tổng quan sát).

5. Đánh giá và lựa chọn mô hình

Có rất nhiều tiêu chí để đánh giá liệu mô hình nào sẽ hoạt động hiệu quả nhất đối với một bộ dữ liệu cụ thể do vậy ta sẽ sử dụng các phương pháp khác nhau để có được một mô hình phù hợp nhất với bộ dữ liệu của chúng ta.

5.1. Ma trận nhầm lẫn

Có rất nhiều tiêu chí để đánh giá liệu mô hình nào sẽ hoạt động hiệu quả nhất đối với một bộ dữ liệu cụ thể và một trong những phương pháp đánh giá phổ biến nhất đó là ma trận nhầm lẫn đã được trình bày ở phần trên với mỗi ma trận tương ứng với một mô hình phân lớp cụ thể. Đối với phương pháp đánh giá này ta thứ ta cần quan tâm chính là tỷ lệ mắc sai lầm của mô hình. Bởi vì đây là những mô hình được xây dựng nhằm mục đích dự đoán do vậy việc có sai lầm là tất yếu, tuy nhiên ta cần không chế tỷ lệ sai lầm của mô hình ở mức thấp nhất có thể vì mỗi một phần trăm sai lầm cũng có thể mang về những thiệt hại rất lớn.

Cụ thể đối với bài toán của chúng ta, điều ta cần chú ý chính là tỷ lệ mắc sai lầm trong việc dự đoán một vị khách sẽ không hủy nhưng thực tế lại hủy, tỷ lệ này cần được không chế ở mức thấp nhất và một mô hình có tỷ lệ này thấp nhất sẽ là mô hình tối ưu nhất. Như đã nói ta cần dự đoán những vị khách nào có khả năng sẽ hủy phòng từ đó tìm cách giữ chân cũng như giảm thiểu những thiệt hại họ gây ra nếu như không thể giữ họ lại. Do vậy việc mắc sai lầm trên sẽ là một điều tai hại vì ta hoàn toàn không thể nhận diện ra họ để tìm cách thuyết phục họ ở lại, đồng thời ta cũng không có bất kỳ sự chuẩn bị nào để giảm thiểu thiệt hại mà họ gây ra khi hủy phòng. Do đó ta cần chọn mô hình với tỷ lệ này càng thấp càng tốt.

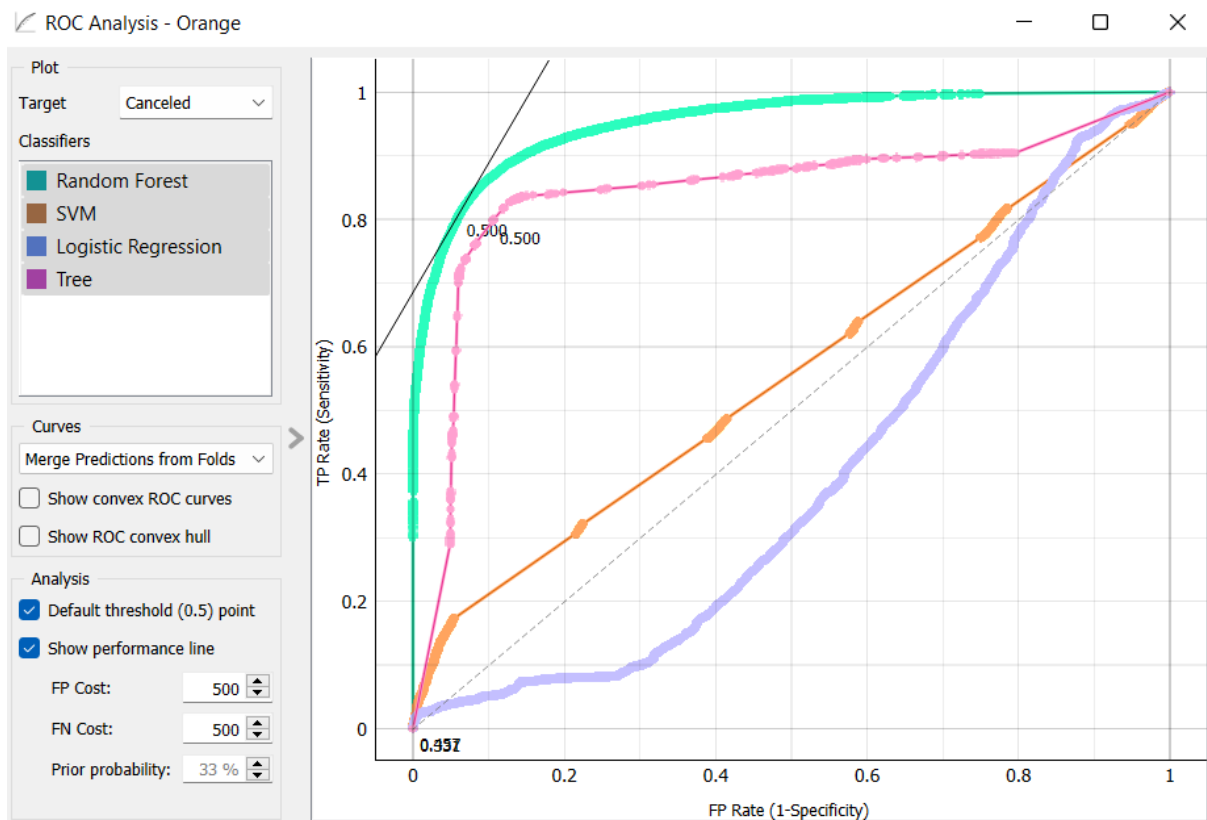
		Predicted		
		Canceled	Not_Canceled	Σ
Actual	Canceled	86.5 %	9.1 %	11885
	Not_Canceled	13.5 %	90.9 %	24390
Σ		11095	25180	36275

⇒ Với phương pháp đánh giá này, Random Forest là mô hình tối ưu nhất đối với bộ dữ liệu này.

5.2. ROC

Đường cong ROC được tạo ra bằng cách biểu diễn tỷ lệ dự báo true positive rate (TPR) dựa trên tỷ lệ dự báo false positive rate (FPR) tại các ngưỡng khác nhau. Qua quan sát đường ROC có thể thấy được mô hình Random Forest là đường có tỷ lệ báo true positive rate (TPR) cao nhất và tỷ lệ false positive rate (FPR) thấp nhất (đường ROC gần điểm có tọa độ (0,1) nhất).

⇒ Phương pháp đánh giá này cho thấy mô hình Random Forest là tối ưu nhất đối với bộ dữ liệu.



5.3. Test and Score

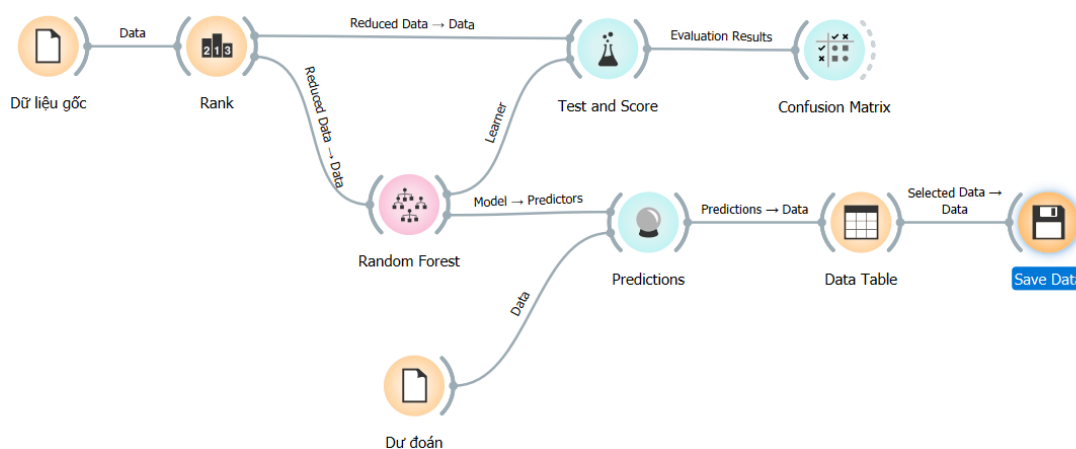
Sử dụng Test and Score widget để xem thêm về các chỉ số AUC (phần diện tích nằm dưới đường cong ROC được trình bày ở phần trên); CA (tính chính xác - tỷ lệ dự đoán chính xác của mô hình với cả 2 lớp); Precision (độ chính xác – tỷ lệ dự đoán chính xác lớp tích cực (positive) trên tổng số dự đoán lớp tích cực); Recall (True Positive Rate – độ phủ - tỷ lệ dự đoán chính xác lớp tích cực trên tổng số thực tế lớp tích cực); F1 (giá trị trung bình điều hòa của hai độ đo Precision và Recall). Các chỉ số này càng cao thì mô hình càng chính xác trong việc phân loại lớp. Tuy vậy thông thường ta sẽ tập trung vào chỉ số CA vì hầu hết mô hình dự đoán đều quan tâm đến tỷ lệ dự đoán chính xác ở cả 2 lớp mà không phải một lớp riêng lẻ bất kỳ.

Model	AUC	CA	F1	Precision	Recall
Random Forest	0.952	0.896	0.895	0.895	0.896
Tree	0.843	0.863	0.863	0.863	0.863
SVM	0.547	0.689	0.646	0.661	0.689
Logistic Regression	0.399	0.672	0.541	0.452	0.672

⇒ Mô hình Random Forest có cả 5 chỉ số cao hơn các mô hình còn lại. Phương pháp Random Forest là phương pháp dự báo tối ưu nhất với bộ dữ liệu của chúng ta.

5.4. Kết luận

Như vậy có thể thấy với bất kỳ phương pháp đánh giá nào thì Random Forest đều là kết quả tối ưu nhất cho thấy đây chính là mô hình thích hợp nhất để áp dụng trong bài toán dự báo này.



6. Giải pháp

Vì những chính sách khuyến khích khách hàng ở lại đều tiêu tốn nguồn lực của khách sạn mà nguồn lực là có hạn do vậy không thể áp dụng tất cả chính sách cho mọi khách hàng. Chính vì vậy nhờ việc tìm ra được nhóm khách hàng có khả năng sẽ hủy phòng bằng mô hình dự đoán ta có sử dụng một lượng nguồn lực thấp nhất để đạt được hiệu quả cao nhất. Như vậy bây giờ ta cần có những chính sách nhằm giữ chân những vị khách này hay ít nhất cũng có thể giảm thiểu thiệt hại nếu họ vẫn quyết định hủy phòng.

6.1. Giữ chân khách hàng

Đầu tiên như đã biết một vị khách có thể hủy với nhiều lý do nhưng thường là 2 lý do chính, một là khách có những chuyện đột xuất hai là khách tìm được nơi ở khác tốt hơn. Tất cả chính sách mà chúng ta đưa ra chỉ có thể phát huy hiệu quả với những vị khách với lý do thứ hai còn với lý do đầu hoàn toàn là tình huống bất khả kháng mà ta không thể kiểm soát được, tuy nhiên những tình huống như vậy thường hiếm hơn do vậy lý do hai thường là đa số.

Trước hết, nếu một vị khách kiếm được một nơi tốt hơn sẽ là bất khả thi nếu ta cứ chăm chú vào việc cạnh tranh về giá cả hay chất lượng dịch vụ vì đã không kịp để chuẩn bị rồi. Ta cần nắm bắt được lợi thế của bản thân là những vị khách có nguy cơ hủy phòng này đã đặt phòng ở khách sạn chúng ta rồi do vậy bản thân họ đã có một ít hài lòng với chất lượng khách sạn (thông qua các bài review và phần giới thiệu về khách sạn trên website) của chúng ta. Tận dụng ưu thế này, ta cần thể hiện ra lòng nhiệt thành đối với khách hàng thông qua các hoạt động trước khi đến khách sạn như liên lạc thường xuyên, sẵn sàng trao đổi tư vấn với khách hàng, cho họ có cảm giác hài lòng khi được quan tâm chăm sóc. Như vậy khi khách hàng đã hoàn toàn thỏa mãn về cả chất lượng khách sạn và thái độ phục vụ thì họ sẽ không có động lực để đi tìm hiểu những nơi khác vì họ cảm giác nơi đây đã khiến họ hoàn toàn hài lòng rồi.

Tiếp theo, như đã đề cập một trong những lý do quan trọng khiến tỷ lệ hủy phòng tăng cao chính do sự thuận tiện và dễ dàng trong thao tác đặt phòng trực tuyến. Điều này khiến cho nhiều người chưa cân nhắc kỹ càng đã quyết định đặt phòng và vì dễ dàng do vậy họ sẽ sẵn sàng hủy khi tìm được một nơi khác tốt hơn. Do vậy, để giữ chân những vị khách này ta cần có những ràng buộc nhằm khiến khách hàng cần có trách nhiệm hơn trong việc ra quyết định đặt phòng đồng thời có những chính sách nhằm khuyến khích khách hàng đến như đã hẹn.

Cụ thể cần đưa ra một chính sách hủy đặt phòng (chủ yếu yêu cầu giao trước một phần tiền theo hóa đơn để làm tiền đặt cọc) việc này không chỉ áp dụng cho những khách có khả năng hủy phòng mà nên dành cho tất cả khách hàng ở khách sạn. Một khoản đặt cọc không chỉ là một điều khoản nhằm khiến khách hàng cần cân nhắc trước khi đặt phòng mà còn là một ràng buộc tâm lý đối với khách hàng. Vì dù chỉ là một khoản tiền nhỏ nhưng nó sẽ như một

lời nhắc nhở cho khách hàng khi đang có ý định hủy phòng rằng khoản tiền này có thể mất trắng và hầu hết mọi người đều có một tâm lý không thích mất tiền oan dù số tiền không hề quá lớn. Nhưng nhờ vậy sẽ cản trở họ khi họ có ý định hủy phòng từ đó giúp giữ chân những khách hàng đã đặt phòng.

Tuy nhiên, không phải ai cũng hoàn toàn thoải mái với việc phải trả trước một khoản đặt cọc không phải vì sợ mất tiền mà họ cảm giác mình không nhận được sự tin tưởng từ khách sạn. Do vậy một chính sách ưu đãi đối với những vị khách đến nhận phòng đúng hẹn là một điều cần thiết. Chính sách này không chỉ là tạo động lực để những vị khách có khả năng hủy phòng ở lại mà cũng là để xoa dịu đi tâm trạng của những vị khách đang không thoải mái với khoản tiền đặt cọc. Sự kết hợp của cả hai chính sách nhằm “vừa đấm vừa xoa”, tận dụng tâm lý khách hàng để giữ chân họ mà vẫn không khiến họ có cảm giác khó chịu.

6.2. Giảm thiểu thiệt hại

Tuy đã vận dụng những chính sách để giữ chân khách hàng nhưng chắc chắn những chính sách này không thể giữ chân mọi khách hàng đồng thời như đã nói những chính sách giữ chân không hề có tác dụng với những khách thực sự có chuyện đột xuất. Do vậy ta còn cần những chính sách nhằm giảm thiểu thiệt hại khi những vị khách này vẫn hủy phòng.

Cần nhắc lại một vị khách hủy phòng ngoài việc khiến khách sạn mất đi nguồn thu từ vị khách đó (1) thì còn mất một khoảng nguồn lực dùng để chuẩn bị khi khách tới nhưng lại không đến (2) , ngoài ra còn mất đi cơ hội tiếp đón những vị khách khác đến sau (3). Như vậy một hành động hủy phòng sẽ khiến khách sạn chịu 3 hệ lụy chính, như vậy để giảm thiểu thiệt hại của khách sạn thì ta cần làm loại bỏ đi ít nhất một trong ba hệ lụy này.

Đầu tiên, với những khách hàng có khả năng hủy phòng không cần chuẩn bị phòng quá sớm tránh lãng phí nguồn lực (cách này sẽ hạn chế được hệ lụy (2)), tuy vậy nếu như việc dự đoán không chính xác ta sẽ không kịp chuẩn bị và làm mất hình ảnh khách sạn. Do vậy ta cần có một cách để làm giảm tối đa khoảng thời gian từ lúc đặt phòng đến lúc hủy phòng, nhờ vậy nếu như vị khách đó tới gần sát giờ vẫn chưa hủy thì có thể họ sẽ không hủy phòng và ta vẫn chuẩn bị phòng cho họ như bình thường. Để có thể giảm khoảng thời gian này ta cần quay lại một bước ở chính sách giữ chân đó là thường xuyên liên lạc quan tâm khách hàng. Việc này nếu không thể giữ chân họ lại cũng có thể khiến họ thoải mái và nếu họ có ý định hủy thì sẽ làm nhanh nhất có thể vì họ cũng hiểu được để càng lâu thì thiệt hại gây cho khách sạn sẽ càng lớn. Tuy vậy việc liên lạc phải do những nhân viên đã được huấn luyện bài bản thực hiện để tránh gây phiền cho khách hàng từ đó dễ dàng đuổi khách đi và làm mất hình ảnh khách sạn.

Ngoài ra ta có thể nhận thêm một lượng khách “dự phòng” để thay thế cho lượng khách có khả năng hủy phòng. Việc này có nghĩa ta có thể cho đặt phòng ở một mức cao hơn so với số phòng hiện trống của khách sạn như vậy khi có những vị khách trước thực sự hủy phòng ta sẽ sắp xếp những vị khách mới vào từ đó loại đi được hệ lụy (2) và (3). Tuy nhiên phương pháp này sẽ bị chia ra thành 4 trường hợp có khả năng xảy ra. Đầu tiên lý tưởng nhất là khi

ta dự đoán chính xác vị khách hủy phòng và có một vị khách khác đến sau để thay thế như vậy ta sẽ khắc phục hai hệ lụy nêu trên. Tuy nhiên nếu dự đoán đúng nhưng không có khách đến sau như vậy ta sẽ không chịu hệ lụy (3) đồng thời nếu ta kết hợp với chính sách giảm thiệt hại nêu trước đó ta vẫn có thể khắc phục hệ lụy (2).

Ngoài ra nếu dự đoán sai, khách không hủy cũng không có khách đến sau thì mọi chuyện vẫn hoạt động bình thường. Nhưng nếu khách không hủy mà lại có khách đến sau việc này sẽ rất khó khăn để giải quyết vì dễ dàng làm mất hình ảnh khách sạn, do vậy cần đào tạo nhân viên biết cách ứng phó với tình huống này như đền bù và chuyển khách đến một phòng khác hay chuyển khách đến những khách sạn đối tác khác nếu kín phòng. Dù trong trường hợp này ta cũng không chịu bất kỳ hệ lụy nào từ việc hủy phòng, tuy vậy dù giải quyết tốt cách mấy thì chắc chắn vị khách đến sau sẽ không bao giờ quay lại, do đó để tránh tình huống này xảy ra ta cần đầu tư nhiều hơn để nâng cao độ chính xác của mô hình dự báo.

6.3. Kết luận

Như vậy bằng cách áp dụng mô hình dự báo ta đã có thể tìm ra được nhóm khách hàng có khả năng hủy phòng từ đó áp dụng các chính sách nhằm khai thác nguồn tài nguyên này và biến họ trở thành khách hàng thực sự của chúng ta, từ đó ta có thể nâng cao công suất phòng rồi dẫn đến cải thiện tình hình kinh doanh hiện tại của khách sạn. Hay ít nhất ta cũng có thể giảm đi những thiệt hại do việc hủy phòng gây ra từ đó cũng có thể cải thiện phần nào doanh thu của khách sạn.

Tuy nhiên bất kỳ chính sách nào được đề xuất ở trên cũng chỉ lợi dụng tâm lý của khách hàng để đạt được những kết quả tốt nhất với khách sạn. Do vậy, để có thể thật sự cải thiện tình hình kinh doanh một cách lâu dài và bền vững, ta cần tập trung vào giá trị cốt lõi của khách sạn đó là chất lượng dịch vụ. Bằng những chính sách trên ta chỉ có thể áp dụng lên những vị khách mới biết đến chúng ta lần đầu nhưng nếu họ đã từng sử dụng dịch vụ khách sạn của chúng ta thì chắc chắn rằng những chính sách này sẽ không có tác dụng với họ một lần nữa. Điều này dễ thấy ở tỷ lệ khách hàng quay lại của khách sạn đã được đề cập ở phần II, vì vậy mô hình này có thể giúp khách sạn phát triển tốt hơn nhưng nó không phải yếu tố cốt lõi bởi xét cho cùng mỗi vị khách khi đến đây thứ họ muốn nhận được là một chất lượng dịch vụ tuyệt vời như ăn uống, vui chơi, nghỉ dưỡng, ... chứ không thể chỉ tìm cách để khiến họ đến ở rồi mặc kệ những chuyện khác.

Vậy để có thể phát triển lâu dài, bền vững ta cần nắm được giá trị cốt lõi của doanh nghiệp là gì từ đó cố gắng phấn đấu. phát triển để ngày càng hoàn thiện nó hơn còn những thứ khác chỉ là những công cụ, chúng có thể giúp doanh nghiệp phát triển nhanh hơn, tốt hơn nhưng nó không thể là nền tảng để xây dựng một doanh nghiệp.

7. Phân công

Tên thành viên	Phân công	% Hoàn thành
Nguyễn Quang Nhật	Tìm dữ liệu, nội dung phần 6	100 %
Vũ Nguyễn Thảo Vi	Tìm dữ liệu, nội dung phần 2	100 %
Nguyễn Quốc Việt	Tìm dữ liệu, nội dung phần 1, tổng hợp bài word	100 %
Phan Đình Nhân	Tìm dữ liệu, nội dung phần 3, tổng hợp bài word	100 %
Đình Trọng Hữu	Tìm dữ liệu, nội dung phần 4,5	100 %