



Hotel Profits and Management

Can we predict whether a guest will cancel hotel reservation and eventually maximize profits?

Background Introduction



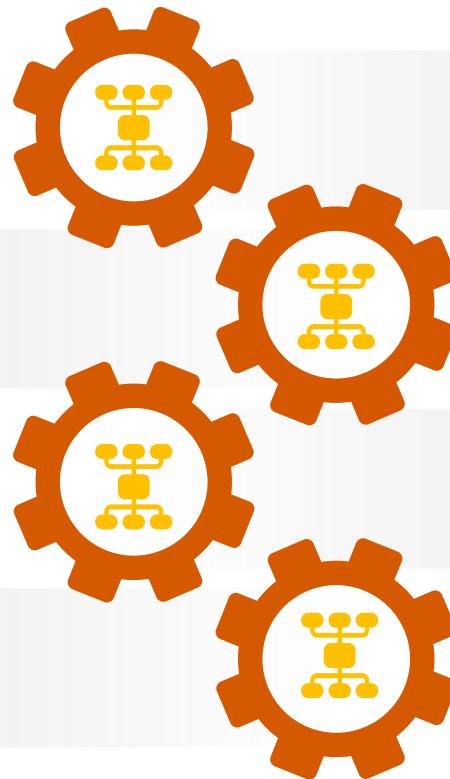
Decision-making Rules

IF profit_Prep >
profit_Not_Prep,

IF Prob (Cancel) > decision cutoff
THEN Not Prepare, ELSE Prepare

ELSE hotel will choose not to
prepare the hotel room

THEN hotel will choose to prepare
the hotel room



Select Appropriate Model - 4 Steps

Probability Prediction

- DV=the probability of hotel-booking that 'is_canceled'
- Narrowed down to 5 models
LR, CT, NN,
k-NN, Random Forest

01



04



02



03



Final Selection

- DV=Probability
- Large dataset
- OKAY-speed of training and generating predictions
- Solid reasoning to interpret
- FINAL 3:
 - LR
 - CT
 - Random Forest

Criteria-Data Volume and Performance

- Large number of observation (n=119,390)
- Small number of missing values
- Avoid:
 - kNN

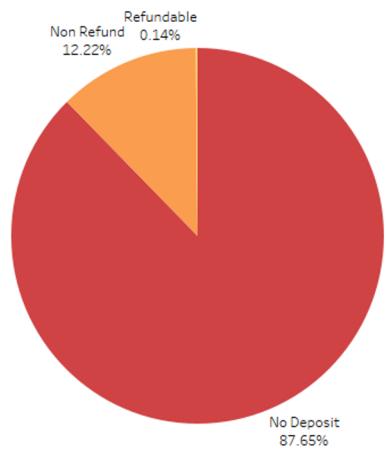
Criteria-Speed and Interpretability

- Slowest speed of predicting result---
kNN
- Need solid reasoning to present and explain
- Avoid:
 - kNN
 - NN

Exploration and Variable Transformation

Variable: deposit

'deposit_type'
(categorical)



'no_deposit'
(categorical)

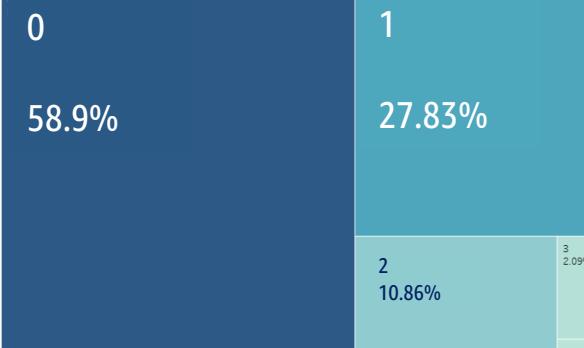
Variable: special request

'Special_request'
(float64)

'make_special_requests'
(categorical)

Two main categories

- 87.65% of the hotel-bookings are with no deposit
- 12.22% of the hotel-bookings are not refundable

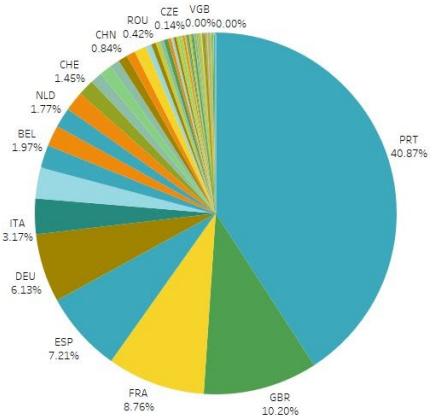


More than 50% of hotel-] bookings do not have special requests

Exploration and Variable Transformation

Variable: country

'country'
(categorical)

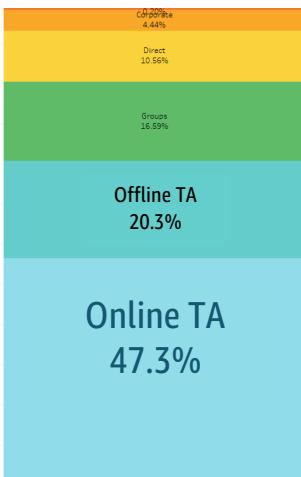


'origin_Portugal'
'origin_other_countries'
(categorical)

- Consists of 178 countries
- 40.87% of the consumers are from Portugal

Variable: market segment

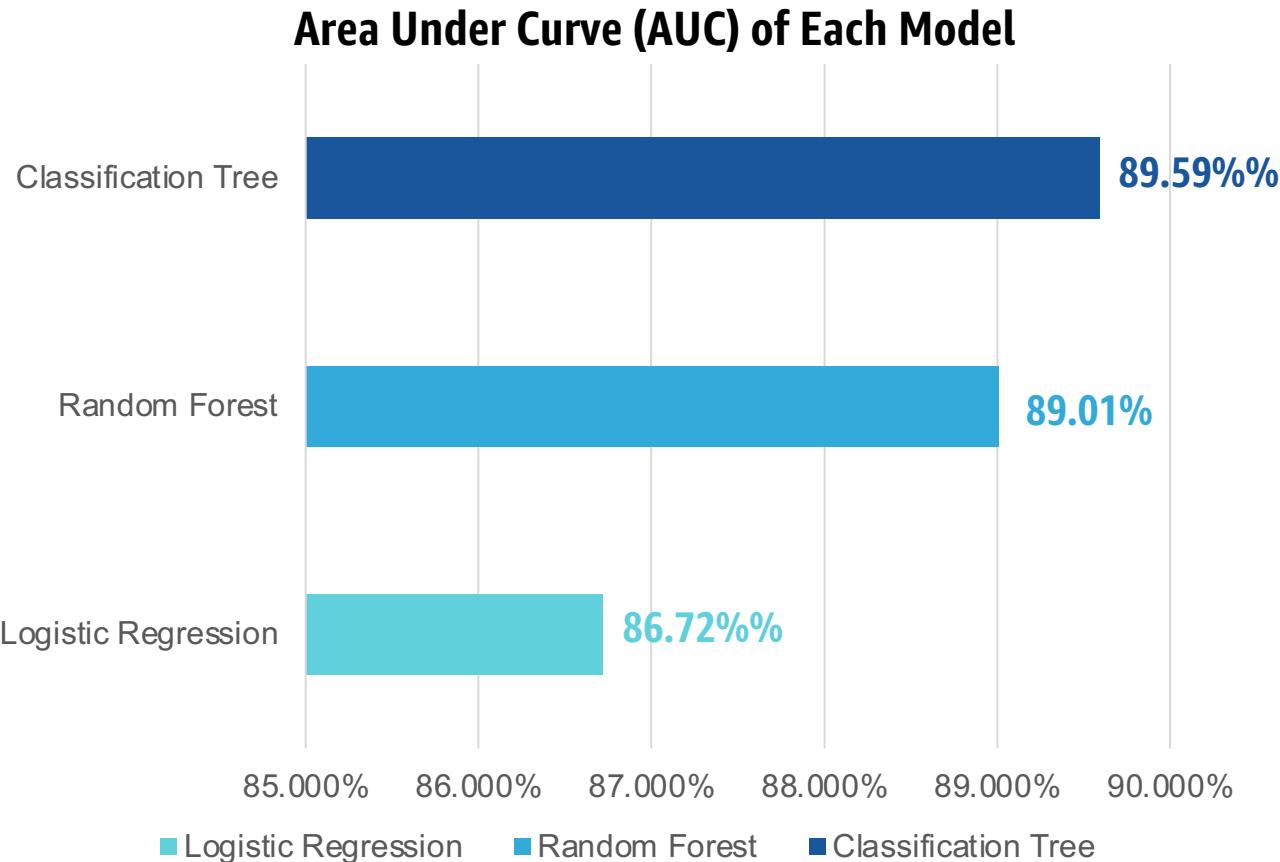
'market segment'
(categorical)



'online_purchase'
'not_online_purchase'
(categorical)

- nearly 50% (47.3%) of hotel-booking are made through Online Travel Agency
- 20% of hotel-booking are made through Offline Travel Agency

Results of 3 models



Modeling-Classification Tree Prediction

The level of depth

10

The number of leaf nodes

390

TOP 3 English Rules

Attached in the speaker
notes at the bottom

Common patterns among the top
leaf nodes

If a customer reserved a hotel room

When the customer is

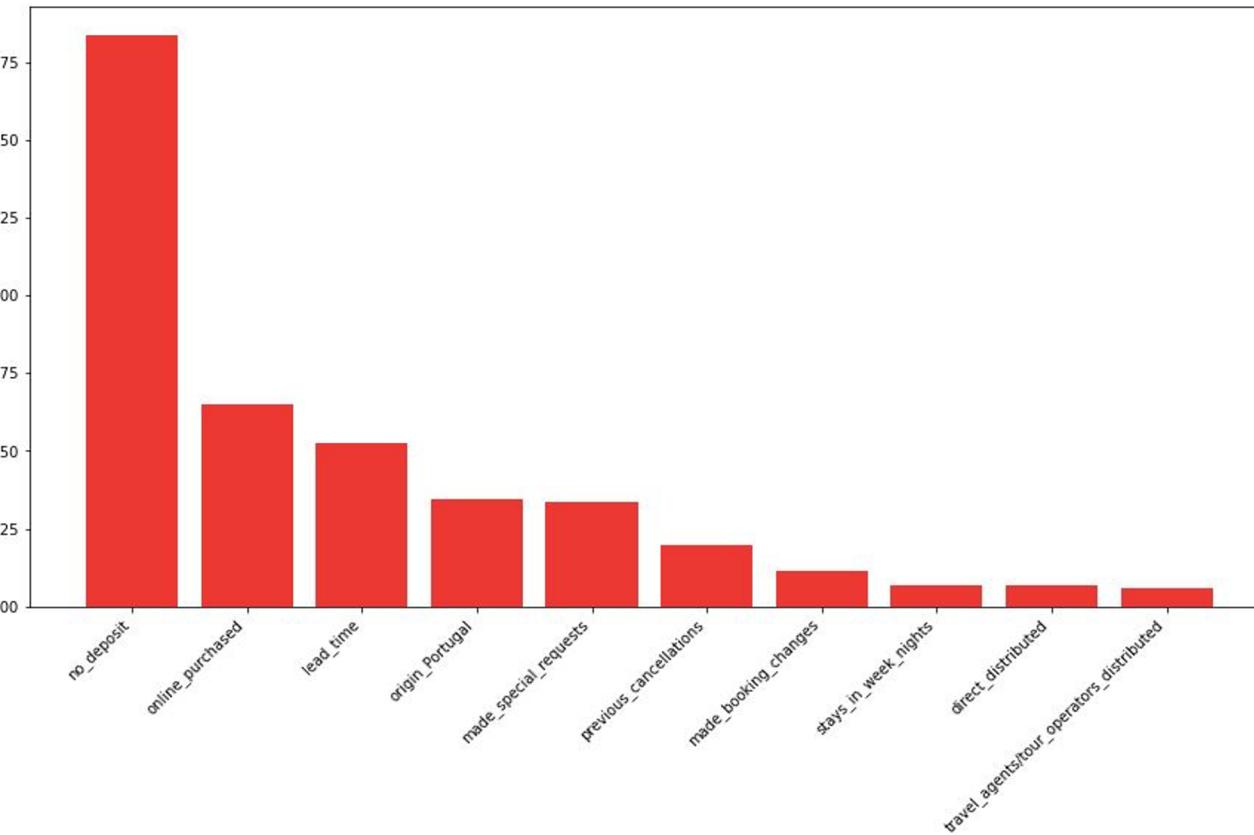
- with children
- with no records of previous
cancelation of hotel-reservation
- a repeated guest to this hotel

Furthermore, when the customer

- made special requests
- reserved 2 nights and more

The customer tend not to cancel the
booking and to show up for check-in

Feature Importance-Classification Tree



- The most important features to predict hotel cancellation are:
- 'No_deposit'
 - 'Online_purchased'
 - 'Lead_time'

Profit Calculation

Hotel Profit Calculation

Customer \ Hotel	Not Prepared	Prepared
Customer		
Not Coming	0	-2
Coming	8	10

Profit Calculation of Each Models

	Average Net Profit	Total Profit
Logistic Regression	\$5.88	\$702392.5
Random Forest	\$5.92	\$706630
Classification Tree	\$5.98	\$714047.5

How Prediction influences decision-making

If the guest has higher predicted probability to cancel the booking, we then will make a decision of not preparing the room for guest, or find ways to overbook.

Net Profit Summary

The best performed model is from **Classification Tree**, which will generate \$5.98 average net profit from each bookings and **\$714047.5 total income** for 3 years.

EVALUATION & SUMMARY

We think that we can very accurately advise our client on whether or not an individual room will be canceled so long as they adopt the **Classification Tree** prediction model.

We can also confidently say that by following the method our hotel clients can maximize both their average and overall profit.

Average Net Profit

\$5.98

Total Net Profit

\$714047.5

Area Under Curve





Thanks.

Data Mining For Business
Team 5
December 2020