



# From Predictions to Prescription

Medical Appointment No-Show prediction and  
overbooking policy prescription



Owen Bao, David Izrailov  
Machine Learning Under a Modern Optimization Lens - 15.095



# Table of contents

**01** Problem/Importance

**02** Data

**03** Methods

**04** Results

**05** Discussion

**06** Conclusion

# Problem/Importance

## Scheduling Disruptions

No-shows leave empty appointment slots, forcing other patients to be treated later. Delayed care can worsen health conditions and increase long-term system costs

## Decreases Physician Utilization

Unused appointment slots reduce provider productivity and create idle time, which is highly inefficient for clinics operating under limited resources

## Lower Quality of Care for All Patients

Missed appointments disrupt clinic flow, causing longer waitlists and less predictable scheduling, which reduces overall patient accessibility and care quality

## Why Predict No Shows?

Predictions can allow us to adjust schedules accordingly, maximizing both the quantity of people receiving care and the quality of that care



# Original Data

The data was collected in Vitoria, Brazil (A city with ~300k residents) and tracked whether or not patients showed up to their appointment.

**n=110,527**

**20% No Show Rate**

Feature	Description	Notes
Patient ID	Identification of a patient	Drop
Appointment ID	Identification of each appointment	Drop
Gender	Male or Female	Make Binary → M=0, F=1
Schedule Day	The day of the actual appointment	Used to Calculate Lead Time
Appointment Day	The day someone booked the appointment	Used to Calculate Lead Time
Age	Age of patient	Drop Age $\leq 0$ & Age $\geq 100$
Neighbourhood	Where the appointment takes place	Drop, not important
Scholarship	If Patient has Financial Aid: 1, Not : 0	Make Categorical and Use
Hypertension	If Patient has Hypertension: 1, Not : 0	Make Categorical and Use
Diabetes	If Patient is Diabetic: 1, Not : 0	Make Categorical and Use
Alcoholism	If Patient is Alcoholic: 1, Not : 0	Make Categorical and Use
Handicap	If Patient is Handicapped: 1, Not : 0	Make Categorical and Use
SMS_received	If Patient received SMS: 1, Not : 0	Make Categorical and Use
No-Show	If Patient is a No-Show: 1, Not : 0	Target Variable

# Data Cont.

## Additional Features

Created: Lead Time

Standardized: Age, Lead Time

Converted: T/F to Binaries to Categorical

## Summary Statistics

Gender Distribution: Female: ~65% Male: ~35%

Average Age: 37, Oldest

10% had financial Aid

32% received SMS before appointment

Mostly imbalanced in Binary Categories



## Correlations

no\_show\_binary correlations range from  $-0.06$  to  $0.18$   
no singular feature linearly predicts

Lead time had largest correlation  
The longer people wait, the more likely they are to no-show

Health factors (hypertension, alcoholism, diabetes) had little to no correlation

Age had negative correlation  
Younger people more likely to no show



# Methods - Prediction

We used three methods to predict no-shows in IAI  
Using AUC and Accuracy on the test set as metrics  
Given large dataset and long computation time, we did not use a large grid search  
With more time we would explore more parameters



## RandomForest

Max Depth: 15  
Trees: 200  
Minbucket: 100

**Test AUC: 0.7331**  
Test Accuracy: 0.7978

## CART

Max Depth: 12  
Minbucket: 200

Test AUC: 0.724  
Test Accuracy: 0.7981

## OCT

Max Depth: 6  
Minbucket: 200

Test AUC: 0.7253  
Test Accuracy: 0.7981

# Methods - Prescriptive

## Formulation

Using each method (RandomForest, CART, OCT), we found its corresponding prescription to minimize the cost function.

## Cost Function

We had to make some assumptions for our costs

1) Cost of **missed appointment**: 150 BRL\*

(Average hourly salary of a doctor in Brazil)

2) Cost of **delay** from overbooking: 75 BRL\*

(Half the cost of missed appointment)

Our logic: Idle doctor is worse than someone having to wait a little longer for their appointment

$$\hat{z}_N(x) \in \arg \min_{z \in Z} \sum_{i=1}^N w_i(x) c(z; y_i)$$

Where  $w_i$  is weights from each model

$y_i$  is show (0) or no-show (1)

$z$  is a decision to overbook or not for that slot

$$c(0; y_i) = \begin{cases} 150, & y_i = 1 \\ 0, & y_i = 0 \end{cases}$$
$$c(1; y_i) = \begin{cases} 0, & y_i = 1 \\ 75, & y_i = 0 \end{cases}$$

\*5.31 BRL = 1 USD

# Results With Prescriptions

## RandomForest

### Costs:

Baseline: 30.32  
Policy: 28.44  
Savings: 1.88  
Percentage: 6.20%

### Extra Patients Seen:

0.0388  
4.86% Increase

## CART

### Costs:

Baseline: 30.35  
Policy: 28.63  
Savings: 1.72  
Percentage: 5.67%

### Extra Patients Seen:

0.0409  
5.13% Increase

## OCT

### Costs:

Baseline: 30.24  
Policy: 28.06  
**Savings: 2.18**  
**Percentage: 7.21%**

### Extra Patients Seen:

**0.0444**  
**5.56% Increase**

**OCT** performed the best out of the 3 models  
on both dimensions





# Discussion



## Model Performance

OCT performed the best on both metrics  
RandomForest achieved strong cost reductions,  
but saw the fewest additional patients  
CART overbooked more, at the expense of  
higher overall cost

## Predicted Accuracy

Results show that the shape and calibration of  
predicted probabilities strongly influence  
prescriptive performance, not just predictive  
accuracy.

## Reduce Operational Costs



Every prescriptive policy outperformed the  
baseline, reducing expected costs, increasing  
the number of patients seen



## Aggregating Results

Savings and Patients seen are small per appointment

If a city's clinics handles 110,527 appointments/year  
Savings:  $110,527 * 2.18 =$  **240,798 BRL (~44k USD)**  
Additional Patients Seen: **4907**



# Conclusion



## Takeaways

Integrating predictive modeling with prescriptive optimization **reduces appointment no-show costs** and **improves clinic throughput** in a universal healthcare setting.

**OCT produces the most effective policy**, showing the lowest cost and highest increase in patients seen; highlighting that **not just predictive accuracy** is crucial for prescriptive performance.



## Real Impact

Strategic, data-driven overbooking is beneficial in universal healthcare systems where physician idle time is more costly than patient waiting, supporting more efficient and scalable appointment scheduling.



# Thanks!