Prediction to Prescription – Medical Appointment No-Show Prediction and Overbooking Policy Prescription

Owen Bao, David Izrailov
15.095 Machine Learning Under a Modern Optimization Lens
December 15, 2025

Problem

We aim to solve the issue regarding no-shows for medical appointments. For each person who misses their appointment, a doctor must idle until her next appointment. In addition, the patient who did not arrive precludes another patient from being treated during that time slot. For this reason, we predicted no-shows, which alone can be very helpful in designing clinical schedules.

In addition, we add a prescriptive component where we offer an overbooking policy to solve the two issues associated with no-shows. The prescription is designed to balance the benefit of reducing idle provider time against the risk that too much overbooking can increase patient waiting and degrade service quality.

We are aware of the risks that come with overbooking, which is why an optimization approach is used to ensure the best results. A solution will help reduce the operational and care-quality losses caused by medical appointment no-shows.

Data

The Data we used came from Kaggle and can be found here. The dataset contains information about 110,527 appointments in Vitória, Brazil. Vitória is the capital city of Espírito Santo state in southeast Brazil and is home to around 300,000 residents. The target variable being whether a patient attended their appointment (show) or missed it (no-show).

Table 1: Description of Features in the Medical Appointment No-Show Dataset

Feature	Description
Patient ID	Identification of a patient
Appointment ID	Identification of each appointment
Gender	Male or Female
Schedule Day	The day of the actual appointment
Appointment Day	The day someone booked the appointment
Age	Age of patient
Neighbourhood	Where the appointment takes place
Scholarship	If Patient has Financial Aid: 1, Not : 0
Hypertension	If Patient has Hypertension: 1, Not : 0
Diabetes	If Patient is Diabetic: 1, Not : 0
Alcoholism	If Patient is Alcoholic: 1, Not : 0
Handicap	If Patient is Handicapped: 1, Not : 0
SMS_received	If Patient received SMS: 1, Not : 0
No-Show	If Patient is a No-Show: 1, Not : 0

Table 2: Summary Statistics of the Medical Appointment No-Show Dataset

Feature	Description
Age	Range of –1 to 115, Mean was 37.1
No-shows	20.19% (22319) No show, 79.81 (88208) Show
Gender	65% Female, 35% Male
Scholarship	90.17% No, 9.83% Yes
Hypertension	80.28% No, 19.72% Yes
Diabetes	92.81% No, 7.19% Yes
Alcoholism	96.96% No, 3.04% Yes
Handicap	97.97% No, 2.03% (One of the 4 levels of disability)
SMS_received	67.9% No, 32.1% Yes

We also dropped identifiers like Patient ID, Appointment ID and avoided using Neighbourhood due to high cardinality and limited interpretability in our setting.

We retained demographic and health indicator variables and treated them as categorical/binary inputs. To aid in our prediction, we added a column for lead time (date scheduled – appointment date) as we think this is relevant to no-show probability. In addition, we standardized our two numeric features of age and lead time.

We cleaned the data by removing extreme outlier ages as well as negative lead times. We also removed certain data from certain neighbourhoods where there were less than 55 appointments in the whole dataset (only 5 neighbourhoods). So after all of this, we were left with 110,458 appointments in our dataset.

Exploratory Findings

To assess linear relationships between features and appointment no-shows, we computed a correlation matrix across numerical and binary variables. Overall, correlations with the no-show indicator are not very high, indicating that no single feature strongly explains attendance behaviour on its own. This is fairly reasonable because no-show can be impacted by real life unpredictable circumstances.

The reason we made lead-time is that we found that it actually exhibited the largest positive correlation with no-shows ($\rho = 0.186$), suggesting that appointments scheduled further in advance are more likely to be missed. SMS receipt is also positively correlated with no-shows ($\rho = 0.126$). Age is negatively correlated with no-shows ($\rho = -0.060$), indicating that younger patients are slightly more likely to miss appointments.

Health-related variables such as hypertension, diabetes, alcoholism, and handicap display near-zero correlations with no-show behaviour. While some of these features are strongly correlated with age (e.g., hypertension–age ρ = 0.505, diabetes–age ρ = 0.292), meaning that they provide very little predictive power for a no-show once age is accounted for.

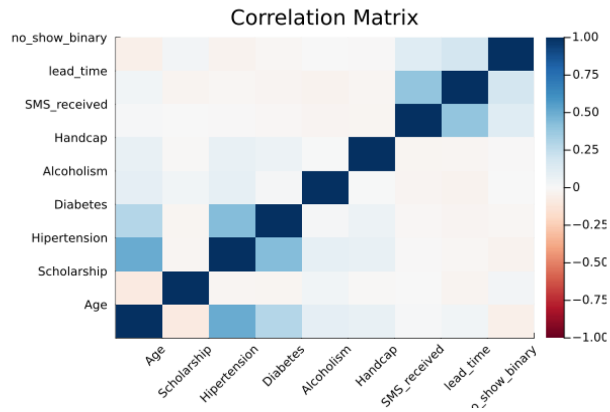

Figure 1: Correlation matrix of features and no-show indicator.

In terms of categorical association, we also examined the relationship between neighbourhood and no show behaviour using a chi-squared test and Cramér's V. The chi-square initially indicated strong statistical dependence ($\chi^2 = 491.9$, $p < 10^{-50}$), however the Cramér's of 0.067 suggested that the association was very weak in practical terms. From this and how there are 81 neighbourhoods in the dataset, it would be very difficult to extract any useful information or predictive value. This is why we excluded it from the primary modelling pipeline.

Using the Optimal Feature Selection (OFS) classifier from IAI, which jointly selects features and fits a linear model through optimization, we examined validation performance as a function of sparsity. The results show that lead time is selected early and drives most of the predictive power, with additional features providing only marginal improvements.

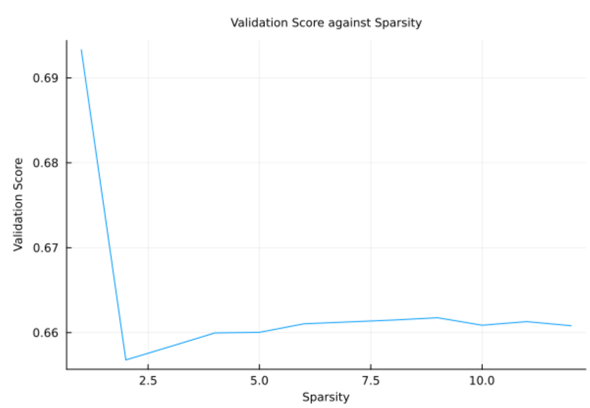

Figure 2: OFS Validation Score versus Feature Sparsity

These findings highlight two key implications. First the absence of strong linear relationships suggests that no-show behaviour is driven by nonlinear interactions rather than a couple dominant linear predictors, thus motivating the use of tree-based methods. Second, since there are no feature perfectly separates no-shows from shows, prediction uncertainty remains substantial. Rather than making scheduling decisions directly from point predictions, we use predicted no-show probabilities as inputs to a downstream optimization model that determines an overbooking policy under uncertainty.

Approach – Prediction

As a baseline, we fit a logistic regression model to estimate no-show probabilities using standardized age and lead time, along with binary indicators for gender, scholarship status, health conditions, SMS reminders. On the held-out test set, the model achieved an accuracy of 0.797**,** comparable to the majority-class baseline given the roughly 80/20 class imbalance. The estimated coefficients align with exploratory findings that longer lead times increase no-show risk (Coef = +0.339) and SMS_received had the largest with Coefficient with 0.357, while age has a negative effect, indicating that younger patients are more likely to miss appointments (Coef = −0.165), the linear structure of logistic regression limits its ability to capture nonlinear interactions observed in the data, motivating the use of tree-based models for improved probability estimation in subsequent analysis.

Using IAI in Julia, we predict using RandomForest, CART, and OCT. We conducted a rather small grid search given limited compute and long runtimes.

Table 3: Selected Hyperparameters from Grid Search for Tree-Based Models

Parameter	Model		
	RandomForest	CART	OCT
Max Depth	15	12	6
Num Trees	200		
Minbucket	100	200	200

We chose this set of models because they range in terms of predictive power and interpretability, while still being similar and a fundamental model level. Given more time, we would have tried other classifiers including SVM, XGBoost, and Neural networks.

Each model was trained with 3-fold cross validation, and a validation criterion of AUC. We also measured accuracy via misclassification. The overall runtime for the prediction component is just over an hour, with OCT alone taking about 40 minutes.

If we were to have more time or more compute power, we could have created a larger grid for a more expansive search and could've led to slightly more improved results, which in this application results in better quality care and more patients being seen.

Results – Prediction

Table 4: Out-of-Sample Prediction Performance Across Models

Score (Out of Sample)	Model		
	RandomForest	CART	OCT
AUC	**0.7331**	0.724	0.7253
Accuracy	0.7978	0.7981	0.7981

As expected, RandomForest has the highest AUC (the metric we trained models to maximize). Though, the single tree models were very close behind, and come with interpretable rules to understand no-show prediction.

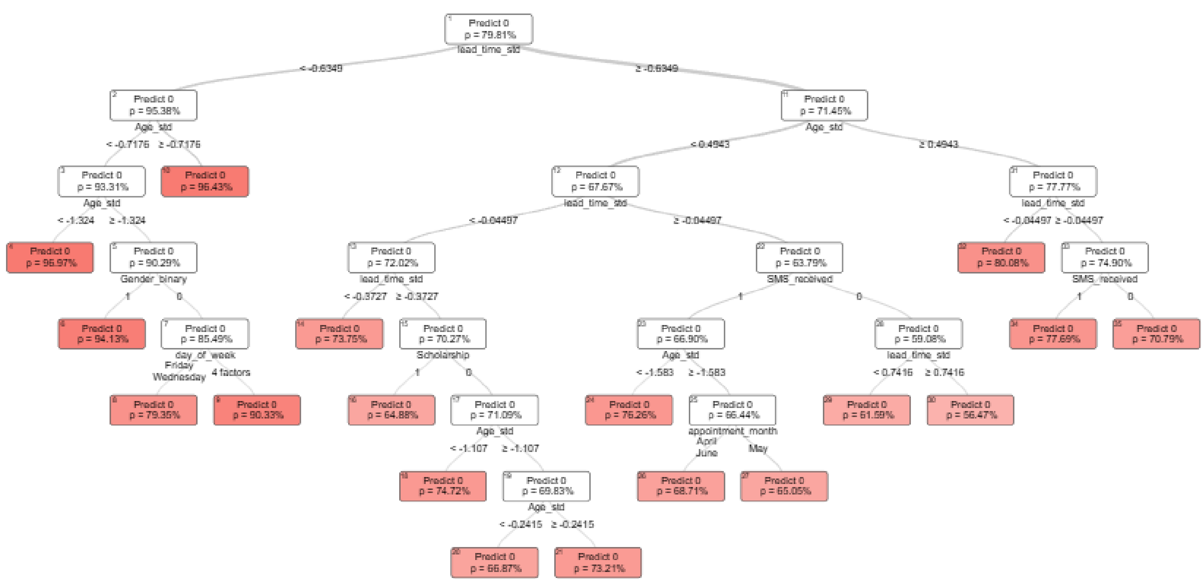

Figure 3: CART Model for Medical Appointment No-Show Prediction

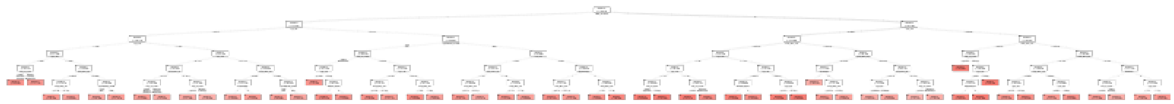

Figure 4: Optimal Classification Tree (OCT) for Medical Appointment No-Show Prediction

Figure 4 shows the Optimal Classification Tree and despite having a smaller depth, it surprisingly looks far less interpretable than the CART in Figure 3.

Approach – Prescription

To facilitate the overbooking policy, we chose to use the prescriptive method framework.

$$\hat{z}_N(x) = \arg \min_{z\in\{0,1\}} \sum_i w_i\,(x)\; c(z; y_i)$$

where $w_i(x)$ is the weights from a model, $c(z; y_i)$ is the cost function, and z is either 1 or 0, where 1 is overbook a given slot and 0 to not.

We chose to try all three models (CART, OCT, RandomForest) as our weights. As for our cost function, we had to make some assumptions. We needed two costs:

1) Cost of idle physician – Cost associated with a missed appointment
2) Cost of patient delay – Cost associated with overbooking when it was not necessary

We believe that the first cost should be higher than the second as less patients are seen under that case. As such, we decided to make cost 1 the hourly wage of a physician in Brazil, around 150 BRL, and cost 2 half that at 75 BRL (note 1 USD ≈ 5.42 BRL). While we understand that these numbers are not the most concrete, we believe that the ratio is about right. We experimented with several ratios for the cost grid and this configuration led to the most reasonable policy.

Table 5: Cost Grid for Overbooking Decision and Appointment Outcomes

Decision	$y_i = 1$ (No show)	$y_i = 0$ (Show)
$z = 0$	150	0
$z = 1$	0	75

This leads to the following cost function:

$$c(z; y_i) = \mathbf{1}\{y_i = 1\}(1 - z)(150) + \mathbf{1}\{y_i = 0\}(z)(75)$$

For OCT and CART we get that:

$$\hat{z}_N^{\text{TREE}}(x) \in \arg \min_{z\in Z} \sum_{i:R(x_i)=R(x)} c(z; y_i)$$

And for RandomForest:

$$\hat{z}_N^{\mathrm{RF}}(x) \in \arg\min_{z \in Z} \sum_{t=1}^{T} \frac{1}{|\{j: R^t(x_j) = R^t(x)\}|} \sum_{i: R^t(x_i) = R^t(x)} c(z; y_i)$$

With this, we are able to prescribe for each appointment whether or not we should overbook a given slot or not, minimizing cost, and maximizing the amount of patients treated.

Results – Prescription

Using Gurobi to solve for these decisions, we can evaluate the baseline costs vs. policy costs. The baseline assumes all patients act as predicted by the model, and the policy is with the optimal prescriptions.

Table 6: Baseline vs. Prescriptive Cost Performance Across Models

Performance	Model		
	RandomForest	CART	OCT
Baseline	30.32	30.35	30.24
Policy	28.44	28.63	28.06
Savings	1.88	1.72	**2.18**
Percentage	6.20%	5.67%	**7.21%**

Table 7: Baseline vs. Prescriptive Patient Throughput Across Models

Performance	Model		
	RandomForest	CART	OCT
Baseline	0.7978	0.7977	0.7984
Policy	0.8367	0.8386	0.8428
Extra	0.0388	0.0409	**0.0444**
Percentage	4.86%	5.13%	**5.56%**

Evidently, the prescription saves a decent amount of costs and also improves patient throughput. Of the three methods, OCT performs the best in both cost saving and patients seen.

Discussion

Our results demonstrate that integrating predictive models with prescriptive optimization can meaningfully improve clinic operations, even when individual prediction accuracy gains are modest. Across all three predictive models, the prescriptive overbooking policy strictly outperforms the baseline of no overbooking, reducing expected costs and increasing patient throughput.

A key finding is that prescriptive performance does not track predictive accuracy perfectly. While RandomForest achieves the highest AUC, it produces smaller throughput gains than CART and OCT. In contrast, OCT consistently delivers the largest cost savings (7.21%) and the greatest increase in patients seen (5.56%). This suggests that the structure of the predictive model can materially influence downstream decisions, even when overall predictive performance is similar across models. In particular, OCT appears to strike a more effective balance between reducing idle physician time and avoiding excessive patient waiting, resulting in superior prescriptive outcomes.

Although the per-appointment improvements are modest, their aggregate impact is substantial. Applying the OCT-based policy to a clinic system handling approximately 110,485 appointments per year yields an estimated savings of 240,798 BRL (approximately $44,430 USD) and enables more than 4,907 additional patients to be treated annually. These results illustrate how small, well-targeted improvements at the individual decision level can scale to meaningful system-wide benefits.

Overall, this project demonstrates the value of a prediction-to-prescription framework for healthcare operations. By combining machine learning–based risk estimation with optimization-driven decision rules, clinics can improve efficiency and patient access without requiring major changes to existing infrastructure. The findings highlight the importance of evaluating models not only on predictive metrics, but also on the quality of the decisions they enable in practice.

Team Contributions:
Owen Bao: 50%
David Izrailov: 50%