Hotel Booking Cancellation Prediction

Chris Katimbwa
11/05/2025

Table of Contents

Business Problem

Hotel cancellations have risen to an average of 40% of overall bookings.

Goal: To accutately predict hotel booking cancellations

- Build a predictive model to forecast cancellations.
- Analyze key factors influencing cancellations
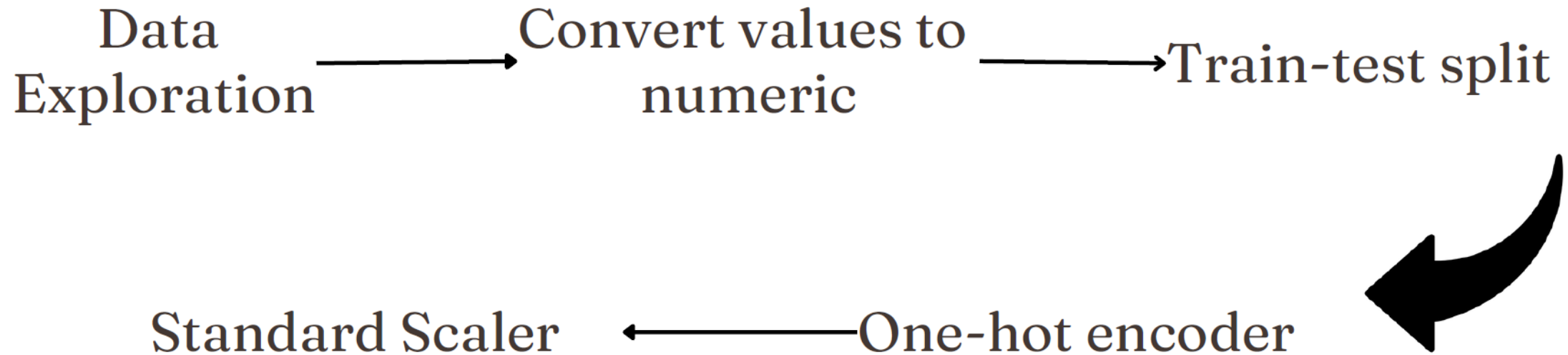- Implement strategies to reduce cancellations.

Data Understanding

Hotel Dataset:
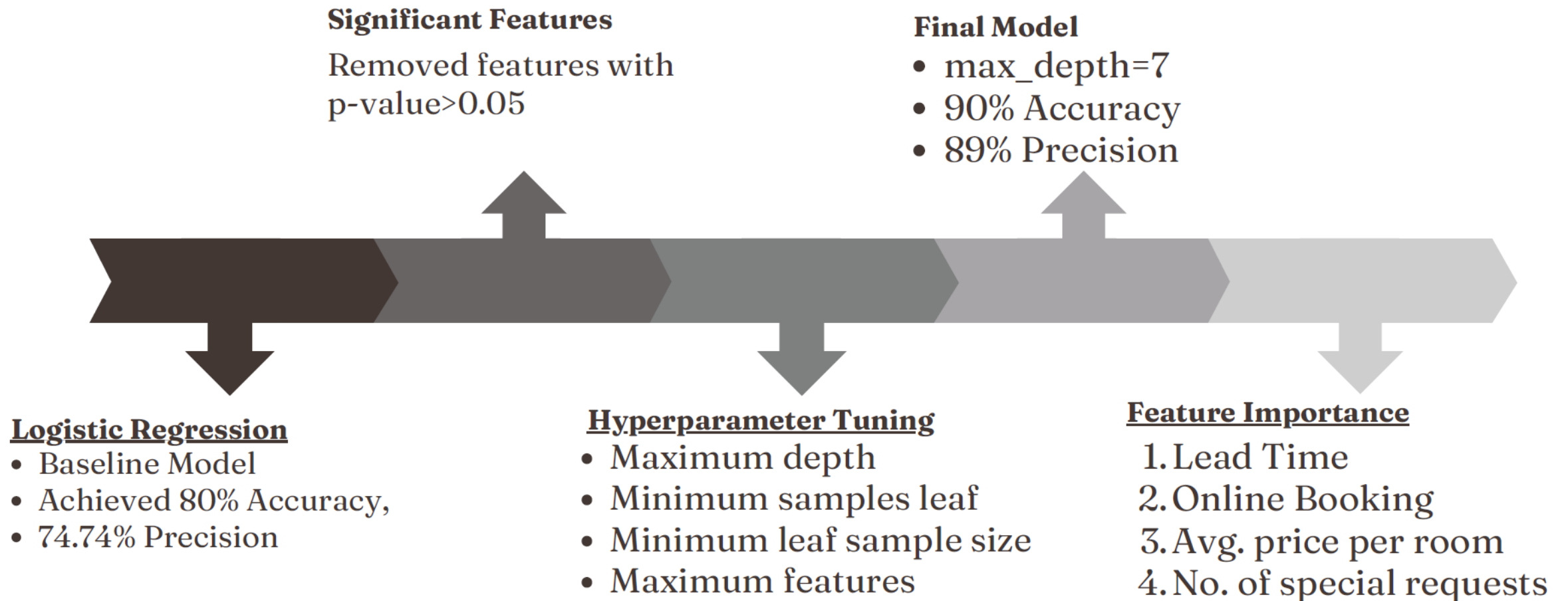- 36275 hotel bookings
- Includes 18 unique attributes of the customer reservation details
- Booking status is our target column.

DATA PREPARATION

Data Exploration → Convert values to numeric → Train-test split

Standard Scaler ← One-hot encoder

MODELING PROCESS

Significant Features

Removed features with p-value>0.05

Final Model
- max_depth=7
- 90% Accuracy
- 89% Precision

Logistic Regression
- Baseline Model
- Achieved 80% Accuracy,
- 74.74% Precision

Hyperparameter Tuning
- Maximum depth
- Minimum samples leaf
- Minimum leaf sample size
- Maximum features

Feature Importance
1. Lead Time
2. Online Booking
3. Avg. price per room
4. No. of special requests
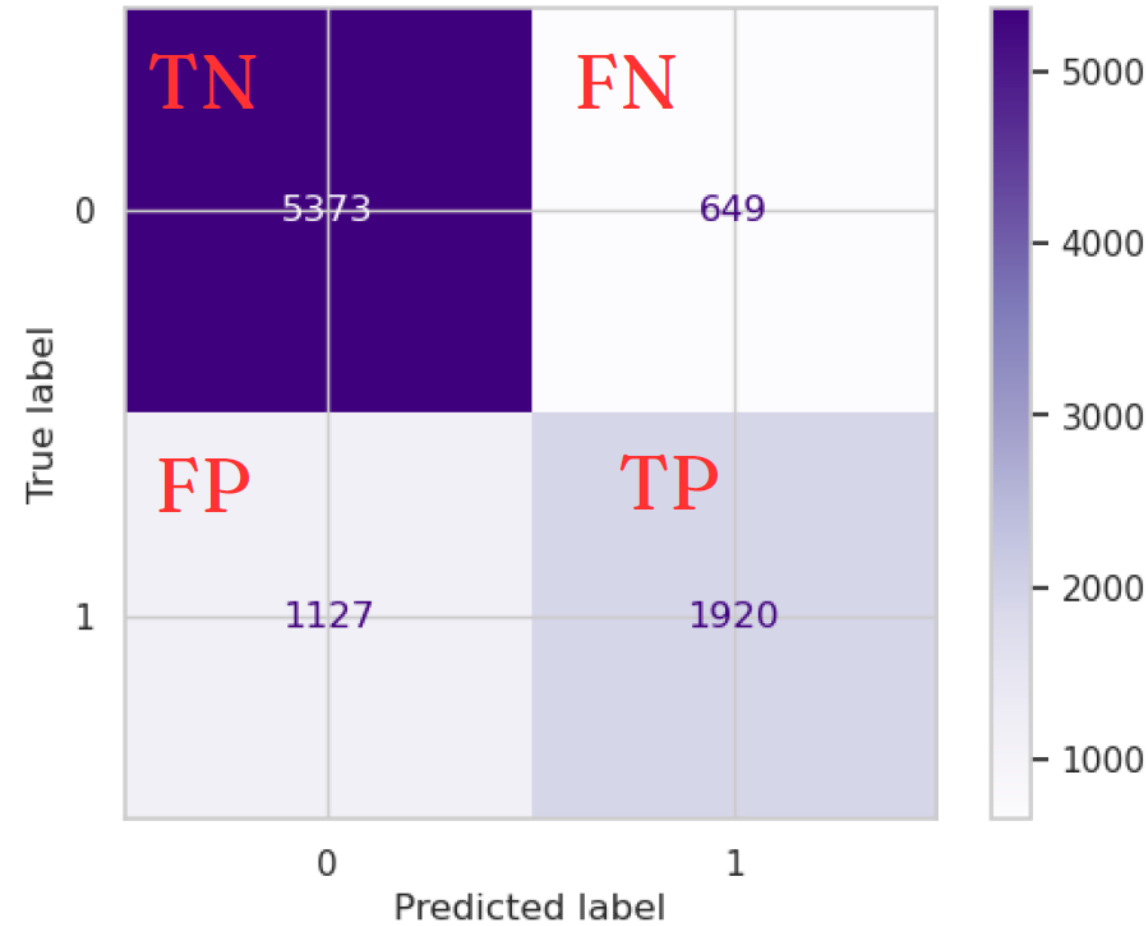
EVALUATION

80% ACCURACY

75% PRECISION

63% RECALL

68% F1 SCORE

0.76 AUC

BASELINE - LOGISTIC REGRESSION

649 retained bookings that were predicted to cancel

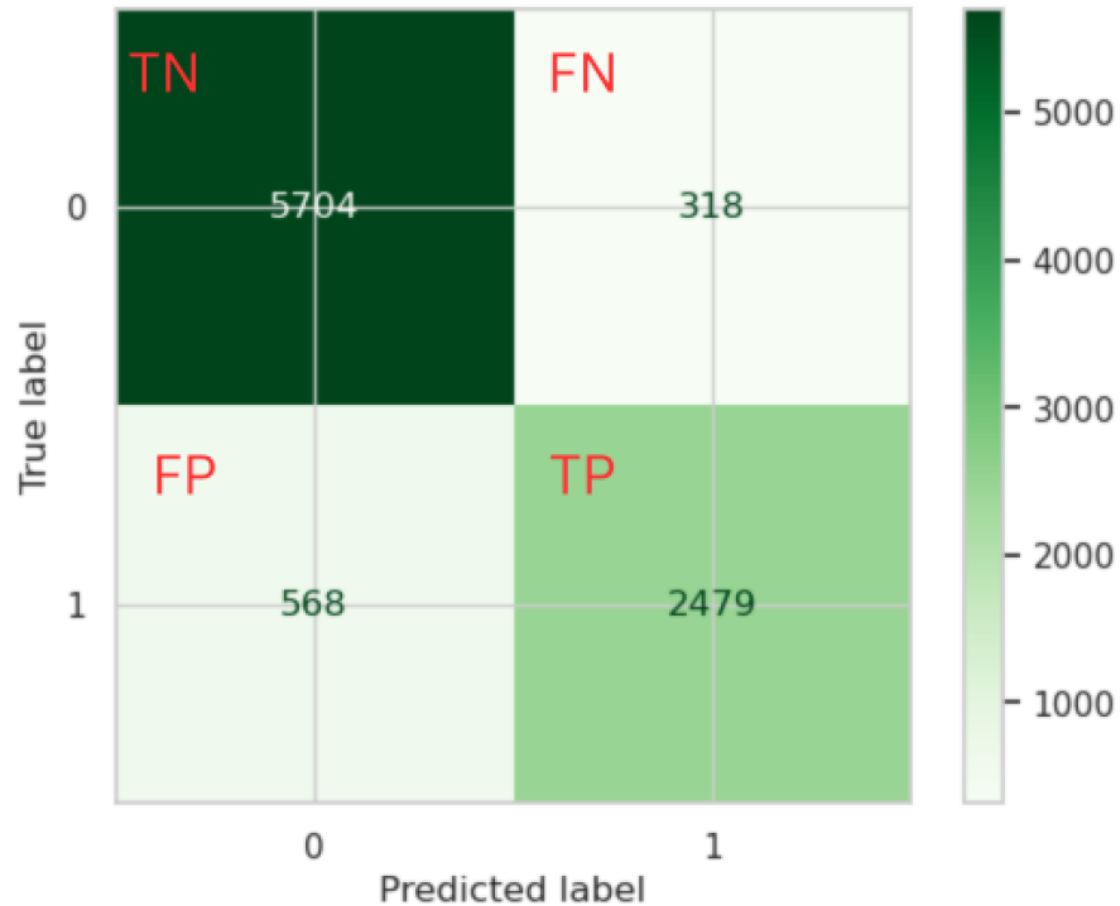# EVALUATION

90% ACCURACY

89% PRECISION

81% RECALL

85% F1 SCORE

0.88 AUC

FINAL - RANDOM FOREST

318 retained bookings that were predicted to cancel

Recommendations

1. Lead time – longer lead times increase cancellation risk. Consider policy adjustments.
2. Dynamic pricing – adjust room rates based on cancellation likelihood.
3. Monitor market segments – tailor strategies to each segment.
4. Loyalty program – offer incentive for booking retention

Next steps

1. Experiment with other ML models to try and improve performance.
2. Gather recent data.
3. Investigate further the top factors of cancellation.
4. Incorporate additional features in the dat like weather and reputation.

THE END