Phase 6: Model Comparison - Logistic Regression vs. CART

Masheia Dzimba and Peter Mangoro

2025-12-06

Contents

1 Introduction

This document presents Phase 6: Model Comparison. We compare the performance of Logistic Regression and CART models, analyze their interpretability, and justify model selection.

2 Load Results

```
Loaded results from both models
```

3 Performance Metrics Comparison

Table 1: Side-by-Side Performance Comparison

Metric	Logistic_Regression	CART	Difference
Accuracy	61.84	60.78	1.07
Precision	64.09	58.71	5.38
Recall (Sensitivity)	42.05	55.83	-13.78
Specificity	79.26	65.16	14.10
F1-Score	50.78	57.23	-6.45
AUC	64.81	60.50	4.31

4 Visualization

4.1 Performance Metrics Comparison

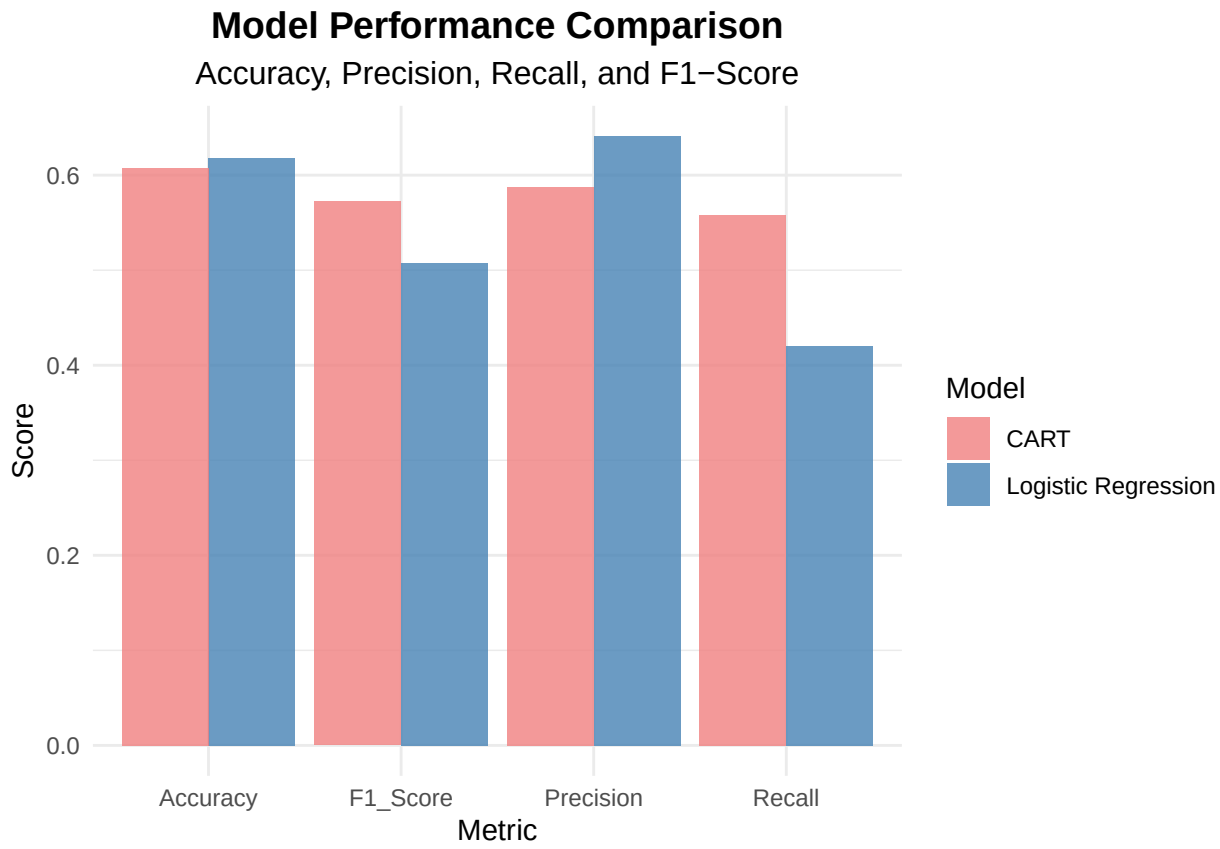

Figure 1: Model Performance Comparison

4.2 AUC Comparison

5 Interpretability Comparison

Logistic Regression:

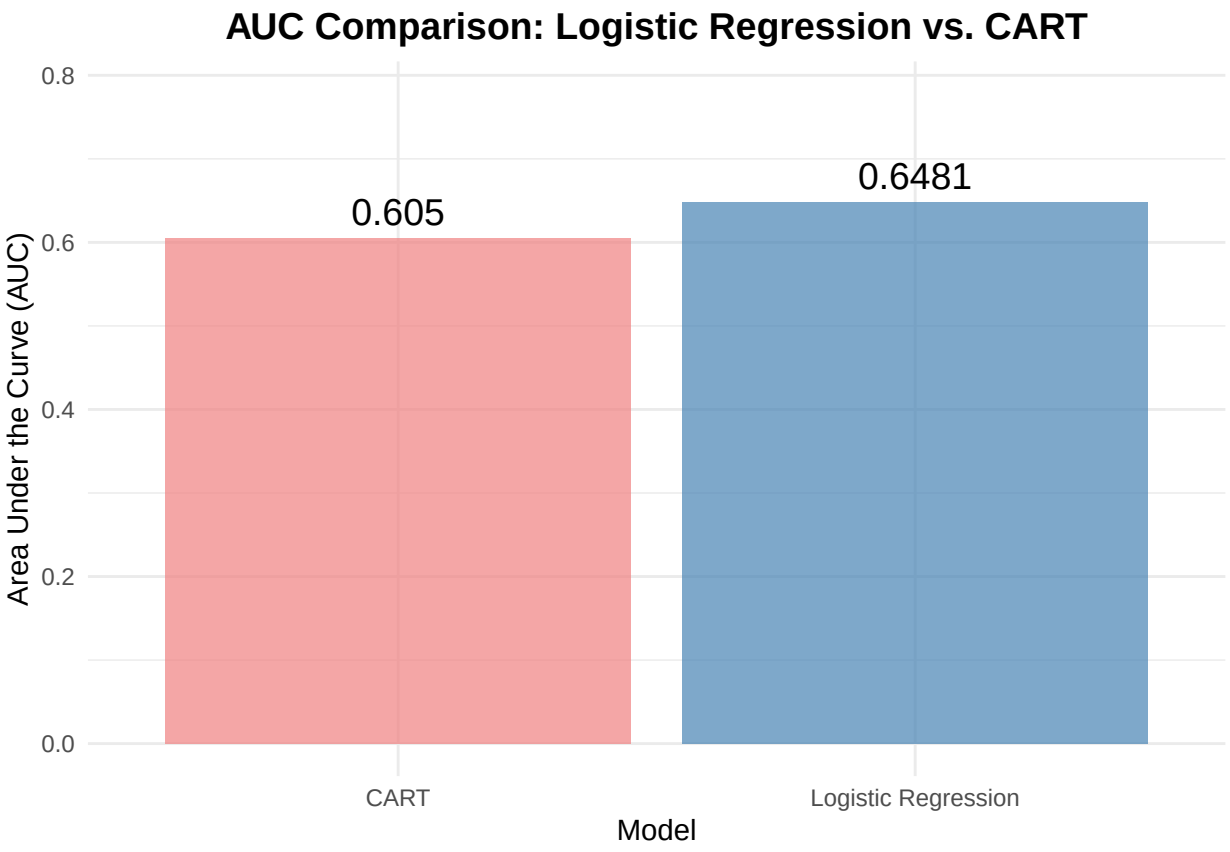

Figure 2: AUC Comparison: Logistic Regression vs. CART

- Provides coefficients and odds ratios for each variable
- Statistical significance testing (p-values)
- 29 parameters in the model
- 17 significant variables ($p < 0.05$)

CART:

- Simple decision tree with 6 variables considered
- Very interpretable: simple decision rule(s)
- Non-linear relationships captured
- Top variable: total_previous_visits (42.32% importance)

6 Model Selection

Performance Metrics Won: Logistic Regression: 4 metrics CART: 2 metrics

Recommended Model: Logistic Regression Reason: Higher AUC and accuracy

Table 2: Model Comparison: Detailed Justification

Criterion	Logistic_Regression	CART
Accuracy	61.84%	60.78%
AUC	64.81%	60.5%
Precision	64.09%	58.71%
Recall	42.05%	55.83%
F1-Score	0.5078	0.5723
Interpretability	High (coefficients, odds ratios)	Very High (simple tree, easy rules)
Complexity	High (33 parameters)	Very Low (1 split, 2 nodes)
Statistical Rigor	High (p-values, hypothesis tests)	Medium (no p-values, variable importance)

7 Key Findings

1. **Performance:**
   - Both models show similar performance (accuracy ~61%)
   - Logistic Regression has higher AUC (0.648 vs 0.605)
   - Both models have fair to poor discrimination (AUC < 0.7)
2. **Interpretability:**
   - CART is simpler (6 variables vs 29 parameters)
   - Logistic Regression provides more detailed statistical insights
   - Both identify previous visits as key predictor
3. **Key Predictors:**
   - Logistic Regression: n_inpatient (OR: 1.47), age groups, medical specialty
   - CART: total_previous_visits (42.32% importance)

8 Summary

This phase compared both models:

- **Logistic Regression** wins 4 out of 6 performance metrics
- **CART** offers superior simplicity and interpretability
- **Recommended**: Logistic Regression (Higher AUC and accuracy)
- Both models identify **total_previous_visits** as the strongest predictor