Phase 5: CART (Classification and Regression Trees) Model

Masheia Dzimba and Peter Mangoro

2025-12-04

Contents

1 Introduction

This document presents Phase 5: CART (Classification and Regression Trees) Model. We build a decision tree model, visualize the tree structure, analyze variable importance, and evaluate model performance.

2 Load Data

```
Dataset:  24996  observations,  18  variables
```

3 Train/Test Split

```
Training set:  17498  observations (70%)

Testing set:  7498  observations (30%)
```

4 Build CART Model

```
# Build model formula
predictor_vars_cart <- setdiff(colnames(data_train_cart), "readmitted")
formula_cart <- as.formula(paste("readmitted ~", paste(predictor_vars_cart, collapse = " + ")))

# Fit the CART model
model_cart <- rpart(formula_cart,
                    data = data_train_cart,
                    method = "class",
                    parms = list(split = "gini"),
                    control = rpart.control(
                      minsplit = 20,
                      minbucket = 7,
                      cp = 0.01,
                      maxdepth = 10,
                      xval = 10
                    ))

print(model_cart)
```

```
n= 17498

node), split, n, loss, yval, (yprob)
      * denotes terminal node

1) root 17498 8227 Not_Readmitted (0.5298320 0.4701680)
  2) total_previous_visits< 0.5 9560 3605 Not_Readmitted (0.6229079 0.3770921) *
  3) total_previous_visits>=0.5 7938 3316 Readmitted (0.4177375 0.5822625) *
```

5 Tree Pruning

```
1-SE Rule CP:  0.01

Pruned tree size:  2  nodes
```

6 Visualize Decision Tree

7 Variable Importance

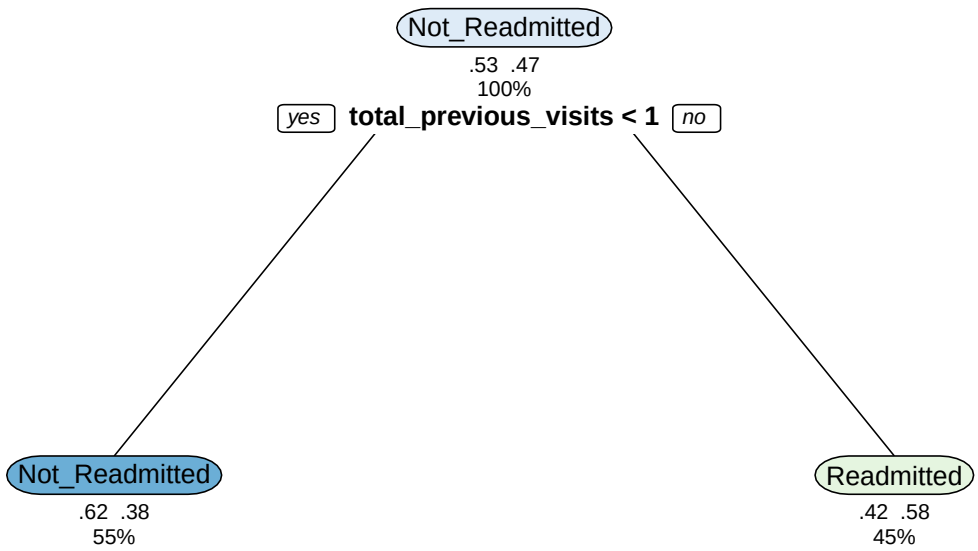
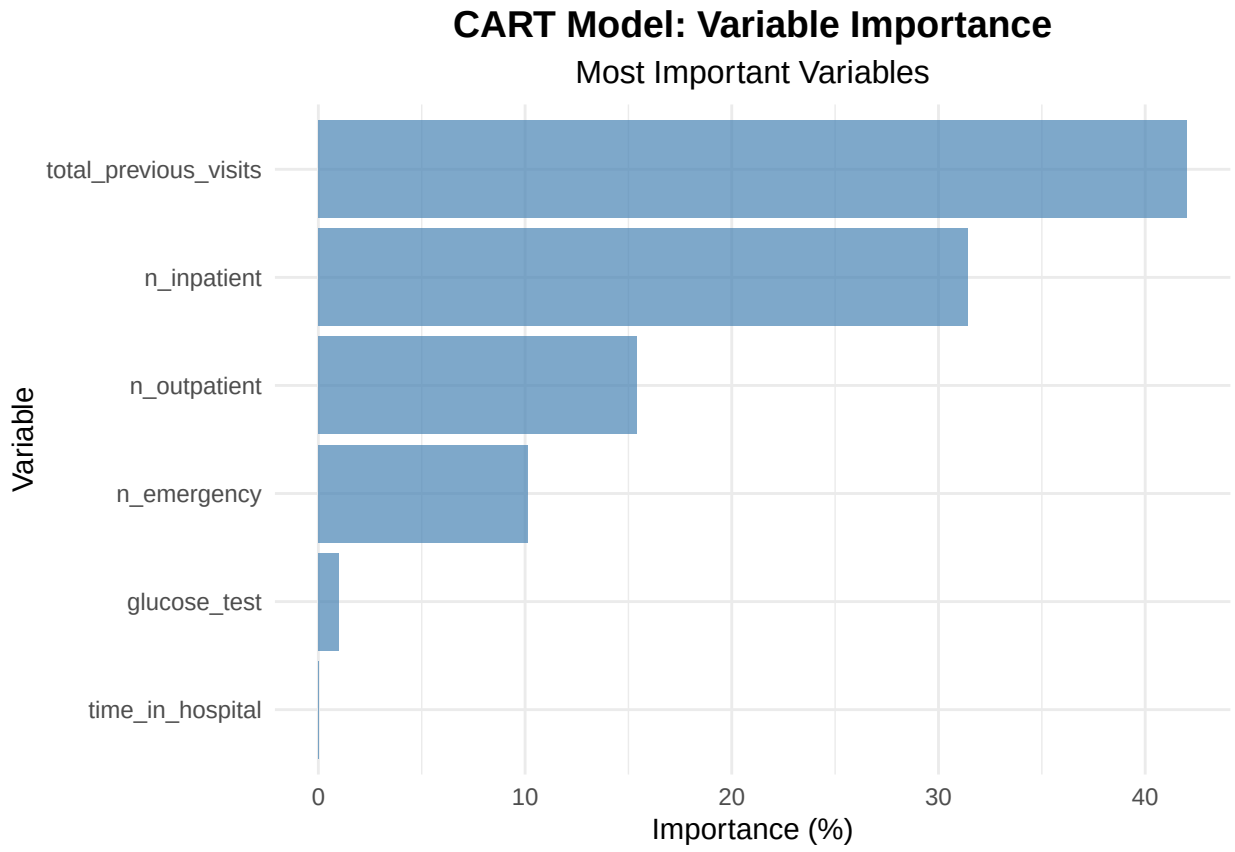


Figure 1: CART Decision Tree: Simple tree with one split based on total previous visits

Table 1: Most Important Variables

Variable	Importance	Importance_Percent
total_previous_visits	365.1239643	42.0200095
n_inpatient	272.9920292	31.4170769
n_outpatient	133.9891796	15.4200413
n_emergency	87.9922076	10.1265153
glucose_test	8.4174459	0.9687153
time_in_hospital	0.4139727	0.0476417



8 Model Evaluation

Table 2: Confusion Matrix

	Not_Readmitted	Readmitted
Not_Readmitted	2601	1529
Readmitted	1372	1996

Table 3: CART Model Performance Metrics

Metric	Value	Percentage
Accuracy	0.61	61.31
Precision	0.59	59.26
Recall (Sensitivity)	0.57	56.62
Specificity	0.65	65.47
F1-Score	0.58	57.91

9 ROC Curve

```
Area Under the Curve (AUC):  0.6105
```

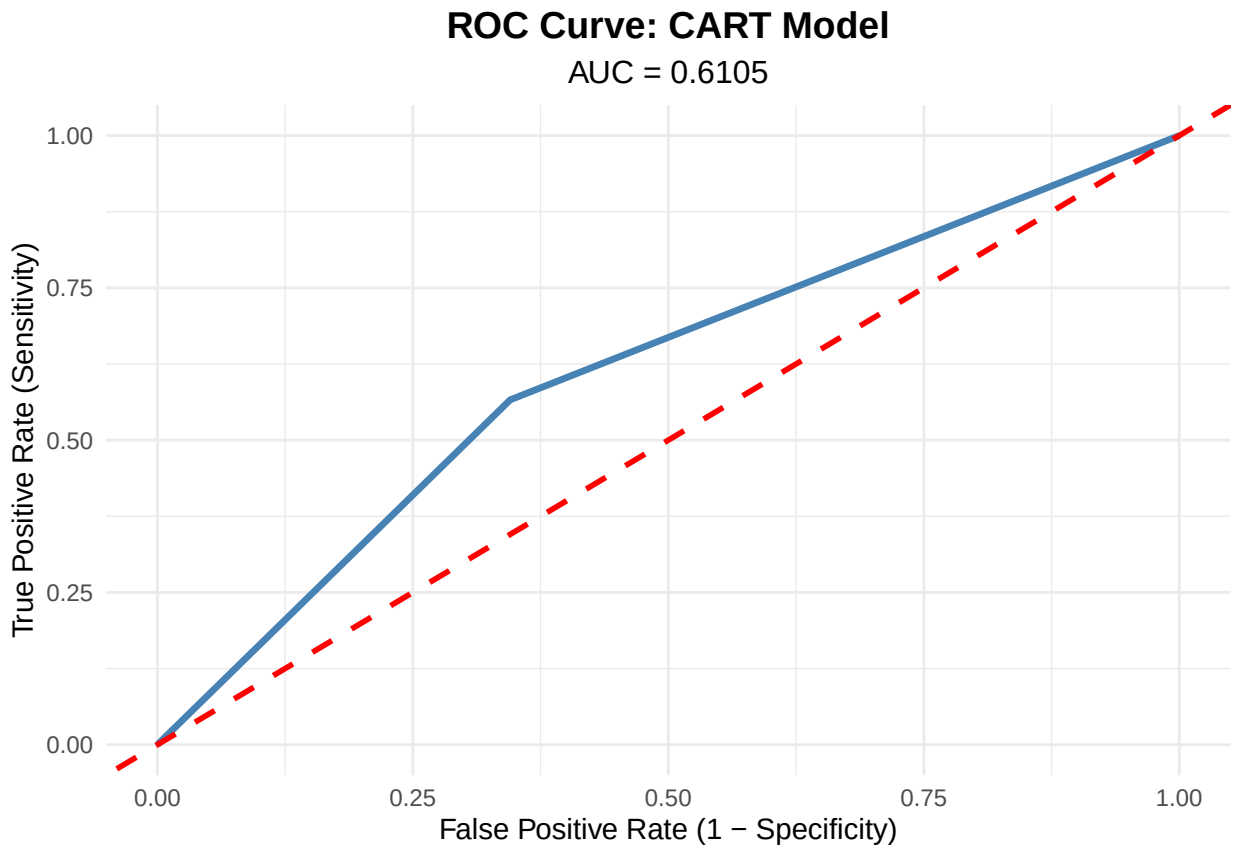


Figure 2: ROC Curve: CART Model

10 Summary

- **Accuracy**: 61.31%
- **AUC**: 0.61
- **Tree complexity**: 1 split, 2 nodes (very simple)
- **Top predictor**: total previous visits (42.02% importance)

- **Interpretability**: Very high (very simple decision rule)