Diagram

Description automatically generated

trainList <- createDataPartition(

y=sub\_2018$MARST,p=.50, list=FALSE)

trainData <- sub\_2018[trainList,]

testData <- sub\_2018[-trainList,]

model.rpart\_smaller <- train(MARST ~ ., method = "rpart",

data = trainData,

tuneLength = 30,

control = rpart.control(maxdepth = 3))

rpart.plot(model.rpart\_smaller$finalModel)

predictValues <- predict(model.rpart\_smaller,

newdata=testData)

confusionMatrix(predictValues, testData$MARST)

Accuracy : 0.735

95% CI : (0.7201, 0.7494)

No Information Rate : 0.4278

P-Value [Acc > NIR] : < 0.00000000000000022

Kappa : 0.5611

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5 Class: 6

Sensitivity 0.9862 0.0000 0.00000 0.23295 0.044335 0.7500

Specificity 0.6130 1.0000 1.00000 0.96881 0.998510 0.9772

Pos Pred Value 0.6557 NaN NaN 0.45055 0.642857 0.9534

Neg Pred Value 0.9835 0.9837 0.98342 0.92002 0.945260 0.8628

Prevalence 0.4278 0.0163 0.01658 0.09893 0.057055 0.3834

Detection Rate 0.4219 0.0000 0.00000 0.02305 0.002530 0.2875

Detection Prevalence 0.6433 0.0000 0.00000 0.05115 0.003935 0.3016

Balanced Accuracy 0.7996 0.5000 0.50000 0.60088 0.521422 0.8636

The model does a good job predicting marital status. We have an accuracy of 73.5%. Compared to a no information rate of 42.8%, this is a drastic improvement. Our sensitivity is 98.6% which is also very good.