# CLASSIFICATION TREE

```{r CLASSIFICATION\_TREE}

set.seed(150)

#Creating 10-fold cross-validation and using down-sampling because of imbalance in data

train\_control\_ct <- trainControl(method = "cv", number = 10, sampling = "down")

#Create sequence of cp parameters to try

grid\_ct <- expand.grid(cp = seq(0.001, 0.3, by = 0.01))

#Train model

ct\_readmission<- train(readmission~., data = train\_data, method = "rpart",trControl = train\_control\_ct, tuneGrid = grid\_ct)

ct\_readmission$bestTune #cp:0.291

ct\_readmission

rpart.plot(ct\_readmission$finalModel)

#variable importance on the final model within training data

varImp(ct\_readmission)

# top three most important variables were age(100), bmi(81.99), weight(65.090) and the least important variable was income.

#accuracy metric and confusion matrix from training.

confusionMatrix(ct\_readmission)

```

# LOGISTIC REGRESSION

```{r}

#LOGISTIC REGRESSION

logistic\_control1 <- trainControl(method = "cv", number = 3, sampling = "down")

set.seed(1000)

logistic <- train(readmission ~ ., data = train\_data, method = "glm", family = "binomial", trControl = logistic\_control1)

summary(logistic)

confusionMatrix(logistic)

confusionMatrix(table((logistic$pred)$pred,(logistic$pred)$obs))