**מדעי הנתונים ובינה עסקית – מעבדה 4**

1. **Commands:**  
wisconsinFile <- read.csv("C:\\Users\\Dan\\Desktop\\wisconsin.csv")  
wisconsin<-na.aggregate(wisconsinFile)  
wisconsin$Class<-factor(wisconsin$Class)

2. **Commands:**  
utils:::menuInstallPkgs() ##for caret and zoo  
library(caret)  
inTrain<-createDataPartition(y=wisconsin$Class,p=0.75 ,list=FALSE)  
training<-wisconsin[inTrain,]  
testing<-wisconsin[-inTrain,]

3. **Commands:**  
NROW(training)

**Results:**   
[1] 524

4. We chose Naïve Bayes:

**Commands:**  
nbFit<-train(Class~ ., data = training,method = "nb")  
nbFit

**Results:**  
Naive Bayes

524 samples

9 predictor

2 classes: '2', '4'

No pre-processing

Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 524, 524, 524, 524, 524, 524, ...

Resampling results across tuning parameters:

usekernel Accuracy Kappa

FALSE 0.9604024 0.9132298

TRUE 0.9600419 0.9116284

Tuning parameter 'fL' was held constant at a value of 0

Tuning parameter 'adjust' was held constant at a value of 1

Accuracy was used to select the optimal model using the largest value.

The final values used for the model were fL = 0, usekernel = FALSE and adjust = 1.

5. **Commands:**  
NbPredict<-predict(nbFit, newdata=testing)  
CM<-confusionMatrix(NbPredict, reference=testing$Class)  
CM

**Results:**  
Confusion Matrix and Statistics

Reference

Prediction 2 4

2 106 0

4 8 60

Accuracy : 0.954

95% CI : (0.9114, 0.9799)

No Information Rate : 0.6552

P-Value [Acc > NIR] : < 2e-16

Kappa : 0.9014

Mcnemar's Test P-Value : 0.01333

Sensitivity : 0.9298

Specificity : 1.0000

Pos Pred Value : 1.0000

Neg Pred Value : 0.8824

Prevalence : 0.6552

Detection Rate : 0.6092

Detection Prevalence : 0.6092

Balanced Accuracy : 0.9649

'Positive' Class : 2