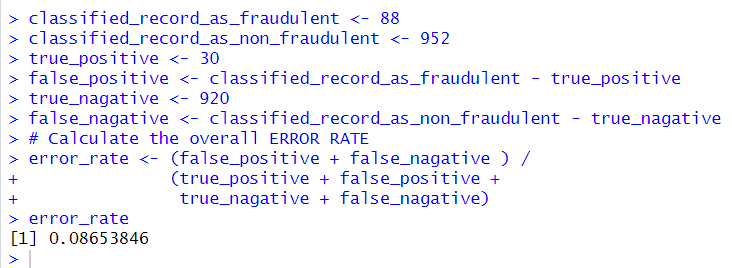
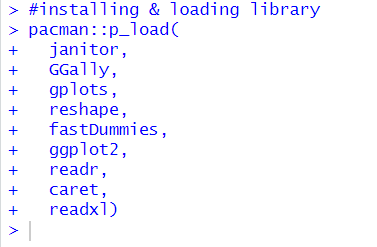
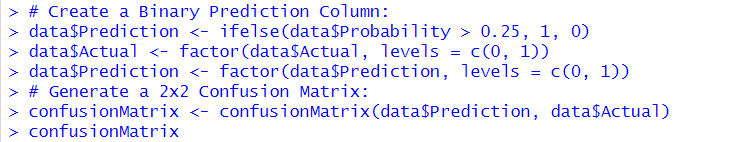
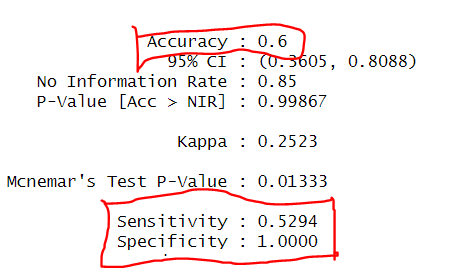
**PART I**

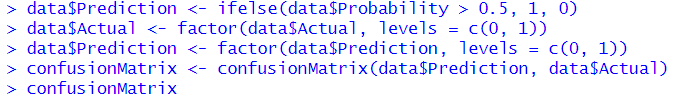
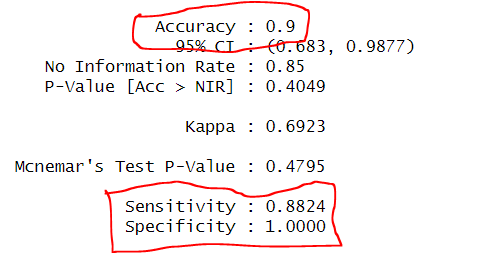
**1.** A data mining routine has been applied to a transaction dataset and has classified 88 records as fraudulent (30 correctly so) and 952 as non-fraudulent (920 correctly so). Construct the confusion matrix and calculate the overall error rate.

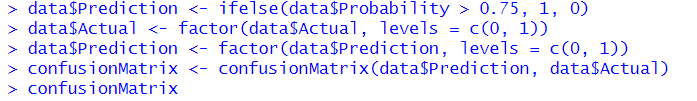
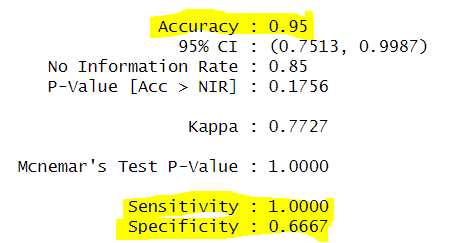


**PART II**

**2.** The following table below shows a small set of predictive model validation results for a classification model, with both actual values and probabilities.  
  
  
  
  
  
  


**a) Create 2x2 confusion matrix for cutoff point = 0.25  
Calculate the overall accuracy, sensitivity, and specificity. Comment on the results.**  
  
  
  
The model's accuracy is moderate at 60%, meaning it correctly predicts an outcome 60 times out of 100, but it's only about 53% effective at identifying true instances. Despite this, it's extremely accurate (100%) at recognizing when an instance is not true, never mistakenly identifying a false instance as true.

**b) Repeat for cutoff point=0.5. Comment on the results**.   
  
  
  
  
This model has a 90% accuracy rate, identifies 88.24% of actual positives correctly, and has a 100% success rate in correctly identifying every actual negative case.

**c) Repeat for cutoff point=0.75. Comment on the results.**  
  
  
  
The model is highly accurate (95%) and perfect at identifying true cases (100% sensitivity), but it's less reliable at ruling out non-cases, with a 66.67% specificity.

**d) Explain carefully what happened to the results when you changed the cut-off point and why.**  
Adjusting the cutoff point in a model is akin to fine-tuning sensitivity; with a low setting (0.25 cutoff), the model is cautious, accurately identifying all true negatives (specificity at 100%) but only captures 60% of the overall scenarios, missing some true positives (sensitivity at 52.94%). As the threshold is increased (to 0.5 and then to 0.75), the model becomes more assertive, correctly identifying more true positives (sensitivity up to 100%) but at the highest setting, it starts mistakenly classifying false signals as true (specificity drops to 66.67%). The challenge lies in finding the perfect balance that maximizes true positive detection without accruing too many false positives, a balance that hinges on prioritizing either comprehensive detection or minimizing incorrect alerts.