Answer Sheet for Assignment: Decision Tree

Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |  |  |
| --- | --- | --- |
|  | **Question** | **Answer** |
| **Part 1. Decision Tree in R**  **(Complexity factor = 0.005)** | | |
| 1 | How often will this tree make a correct prediction (include decimals)? Provide your answer for both the training set and the validation set. | **87.33% for the training set, and 83.67% for the validation set** |
| 2 | How likely is a customer to pay back their loan if they have one child and make $35,000 per year?  *(NOTE: When asked “how likely…” cite the percentage!)* | **94%** |
| 3 | How likely is a customer to pay back their loan if they are married, make $45,000 per year, have no children, and no mortgage? | **14%** |
| 4 | How likely is a customer to pay back their loan if they make $83,000 per year and have no children? | **89%** |
| 5 | Describe the profile of the least likely customer to successfully repay their loan. | **Income less than $51,000; less than 0.5 children (or no child); married; with mortgage; and with saving account**  **(probability = 8%)** |
| 6 | Describe the profile of the most likely customer to successfully repay their loan. | **Income less than $51,000 but greater than $15,000; More than 0.5 children but less than 1.5 children (or with one child)**  **(probability = 94%)** |
| **(Complexity factor = 0.05)** | | |
| 7 | How often will this new tree make a correct prediction (include decimals)? Provide your answer for both the training set and the validation set. | **80% for the training set, and 82% for the validation set** |
| 8 | Is this model better or worse than the first model at predicting who will repay their loan? Explain how changing the complexity factor affected the tree using **no more than two sentences.** | **Worse. The higher complexity factor makes the tree simpler with fewer splits. This makes each node less accurate as it combines different groups.  *(any explanation which is consistent with this is ok)*** |
| 9 | How likely is a customer to pay back their loan if they have one child and make $35,000 per year? | **94%** |
| 10 | Does marriage increase or decrease the likelihood that a customer will pay back their loan? | **Decrease** |
| **Part 2 Compute and Evaluate Decision Trees** | | |
| 11 | What is the correct classification rate for Tree #1? | **Tree #1:**  **(600+1000)/2000 = 0.8**  **Or 1- (230+170)/2000 = 1-0.2=0.8** |
| 12 | What is the correct classification rate for Tree #2? | **Tree #2:**  **(900+800)/2000 = 0.85**  **Or 1- (250+50)/2000 = 1- 0.15 = 0.85** |
| 13 | Which decision tree (Tree #1 versus Tree #2) has higher classification accuracy? | **Tree #2 has higher classification accuracy** |