# **BAN210- Workshop 3**

| Total Mark: | 10 marks |
| --- | --- |
| Submission file(s): | * WS03.docx * *WS3\_pg3-49.lst* * WS3\_pg3-59-Tree.bmp |

Please work in **groups** to complete this lab. This workshop is worth 5% of the total course grade. Please submit the submission file(s) through Blackboard. Only one person must submit for the group and only the last submission will be marked.

## **Part I: Video Tutorial**

1. Watch third video in Getting Started with SAS Enterprise Miner playlist:

* Video (Getting Started with SAS Enterprise Miner: Building Decision Trees): <https://youtu.be/IlUZYlgkeSc>

## **Part II: Data Partitions**

1. Open the project you worked on in Workshop 1. Open the Retention Diagram. If you don’t have access to this project, repeat pages 3-31 to 3-34 of the SAS Advanced Business Analytics course notes (available under BB> Course Information> Resources).
2. *Optional:* Follow instructions on pages 3-46 to 3-49 of the SAS Advanced Business Analytics course notes.
3. Follow steps 9 to 10 on page 3-49 of the SAS Advanced Business Analytics course notes to create the data partitions (training and validation sets). Save the results as *WS3\_pg3-49.lst* and submit with this workshop.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

Chart

Description automatically generated

A picture containing chart

Description automatically generated

Graphical user interface, application

Description automatically generated

## **Part III: Decision Trees**

1. Follow the instructions on pages 3-59 to 3-61 of the SAS Advanced Business Analytics course notes. As you go through the steps, answer the following questions.

Graphical user interface, application

Description automatically generated

Graphical user interface, Word

Description automatically generated

Graphical user interface, chart, line chart

Description automatically generated

Table

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface

Description automatically generated with low confidence

Graphical user interface

Description automatically generated

**Answer the following questions:**

* At step 3 on page 3-59,
  + What is the misclassification rate for the validation set?
  + For the training set, compare sum of squared errors (SSE) with the average squared error (ASE). What is SSE divided by? Is this the same number mentioned as DIV in the Fit Statistics table?
* At step 4 on page 3-59,
  + For the leaves at depth 5, what is the probability of having target = 1? (in training set)
  + From the menu, choose Edit> View> Fit to page.
  + Then save the tree as WS3\_pg3-59-Tree.bmp and submit with this workshop.
* At step 5 on page 3-60, look at the average square error plot. What other value for the *number of leaves* looks promising?
* At step 6 of page 3-60, copy the ‘variable importance’ report and paste here. See page 3-16 and explain which inputs are important for predicting which students will not return to school.
* At step 7 of page 3-60, paste the counts of FN, TN, FP, TP for training and validation sets here. You will be using these numbers in Part IV.
* At step 11 on page 3-61, copy the ‘variable importance’ report and paste here. What is the meaning of the rightmost leaf at level 7 (deepest) in plain English?
* At step 11 of page 3-60, paste the counts of FN, TN, FP, TP for training and validation sets here. You will be using these numbers in Part IV.
* Paste a picture of the retention diagram here.

## **Part IV: Classification Assessment**

1. Use the numbers you obtained above for the validation set by the “Misclassification Tree” and the “Probability Tree” to fill the following tables. Then calculate Precision, Recall, and F1 for both trees

|  |  |  |  |
| --- | --- | --- | --- |
| Misclassification Tree | | | |
|  | Detected as 0  (outcome= 0) | Detected as 1  (outcome = 1) | Total |
| Truly 0  (target = 0) | TN= | FP= | FP+TN = |
| Truly 1  (target = 1) | FN= | TP= | TP+FN = |
| Total | TN+FN= | TP+FP= |  |

Recall (R) = TP / (TP + FN) =

Precision (P) = TP / (TP + FP) =

F1= 2P.R / (P + R) =

|  |  |  |  |
| --- | --- | --- | --- |
| Probability Tree | | | |
|  | Detected as 0  (outcome= 0) | Detected as 1  (outcome = 1) | Total |
| Truly 0  (target = 0) | TN= | FP= | FP+TN = |
| Truly 1  (target = 1) | FN= | TP= | TP+FN = |
| Total | TN+FN= | TP+FP= |  |

Recall (R) = TP / (TP + FN) =

Precision (P) = TP / (TP + FP) =

F1= 2P.R / (P + R) =

## **Part V: Group work**

1. Add this declaration to your file:

We, ------------ (mention your names), declare that the attached assignment is our own work in accordance with the Seneca Academic Policy. We have not copied any part of this assignment, manually or electronically, from any other source including web sites, unless specified as references. We have not distributed our work to other students.

1. Specify what each member has done towards the completion of this work:

|  | Name | Task(s) |
| --- | --- | --- |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |