Group Assignment # 1 Decision Trees and Random Forests

By: Will Singletary, Campbell Jones, Christopher Best, Allana Green

(b)

i.

A screenshot of a computer

AI-generated content may be incorrect.

(h)

1. Minimum(income): 97.0

Maximum(income): 2379.0

Mean(income): 324.9

1. Median(DTI): 0.24
2. Loans denied: 154

Loans approved: 246

1. “insert comments on data”

(k)

ii. Error debugging with GenAI:A screenshot of a computer

AI-generated content may be incorrect.

iii.A diagram of a diagram

AI-generated content may be incorrect.

1. A screenshot of a computer program

   AI-generated content may be incorrect.

1)

1. Employment is among the least important but still plays a visible role in the tree. For those with high income, high credit scores, lower educations and higher DTI, employment plays some role.
2. Gender was used in our rpart(), but did not show up in the decision tree.

2) The importance of classifications, specifically credit score and total income, are shown on the decision tree to be a significant deciding factor, appearing three times. Their importance factors being the highest (>13) are supported visually on the tree by reducing the impurity multiple times.

(I)

iii. Screenshot with title and tweak=1.1

A diagram of a diagram

AI-generated content may be incorrect.

(m)

ii. Tree with “extra=1” argument:A diagram of a graph

AI-generated content may be incorrect.

1. Calculating reduction in impurity:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Count for class 0 | Count for class 1 | Impurity | Weight |
| Root Node | 110 | 170 | 0.6071 | 1.00 |
| Left Child of Root | 29 | 13 | 0.3095 | 0.15 |
| Right Child of Root | 81 | 157 | 0.6597 | 0.85 |

(n) Confusion matrix:

A black background with white text

AI-generated content may be incorrect.

Our model has misclassified 10 approvals as denials as well as misclassifying 29 denials as approvals. It appears that if we were to use our model to determine loan approval, our model would be riskier than the model GTEB has used.

(o) Adjusting model without using credit score:

i. A screenshot of a computer

AI-generated content may be incorrect.

ii. A diagram of a tree

AI-generated content may be incorrect.

iii. A black screen with white text

AI-generated content may be incorrect.

1. When utilizing the credit score, the tree structure appears much more imbalanced versus when using a decision tree without the score. The confusion matrices do not have a significant change, however using credit score correctly predicted 81 results versus 80 correct predictions with education. We consider this an immaterial difference.
2. We have used an iterative R function to help us determine if there is a different loan classification based on the utilization of credit score vs education. There is one record (loan\_id: 2) that was categorized as a denial using credit score and categorized as an approval using education.

(p)

iv. A screenshot of a chat

AI-generated content may be incorrect.

(q) Variable Importance:

A screenshot of a computer screen

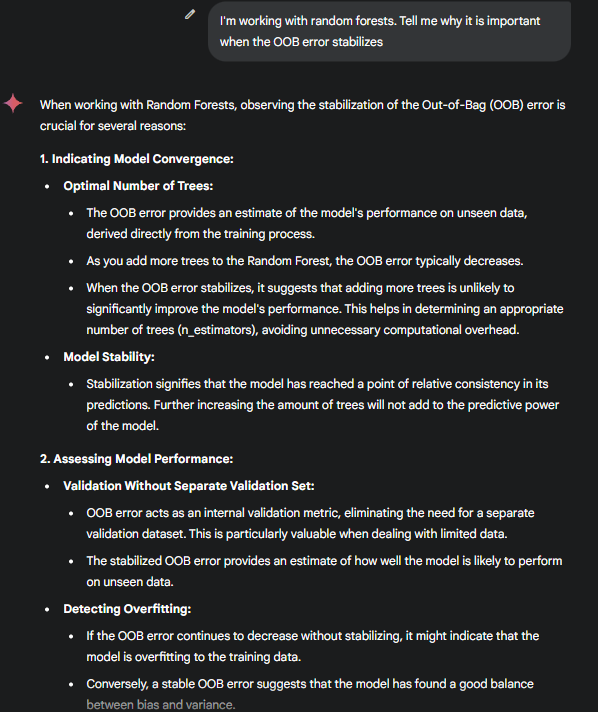
AI-generated content may be incorrect.

The importance of variables seems to have changed significantly, particularly due to the fact we have included total loan value as one of the variables in the random forest. It appears that income is still heavily weighted, but DTI and education have been compressed in their importance, because loan amount holds much significance.

(r)

i. It appears that between 180 and 200 trees are needed in order for the error rate to stabilize.  
A graph of different colored lines

AI-generated content may be incorrect.

ii.  
 

(s) When comparing the confusion matrix of the random forest to the confusion matrix of the decision tree (disregarding credit score), they are very similar with the forest marginally more accurate. The forest has one less false negative and one less false positive. The forest is 68.33% accurate in its predictions and the tree is 66.67% accurate.