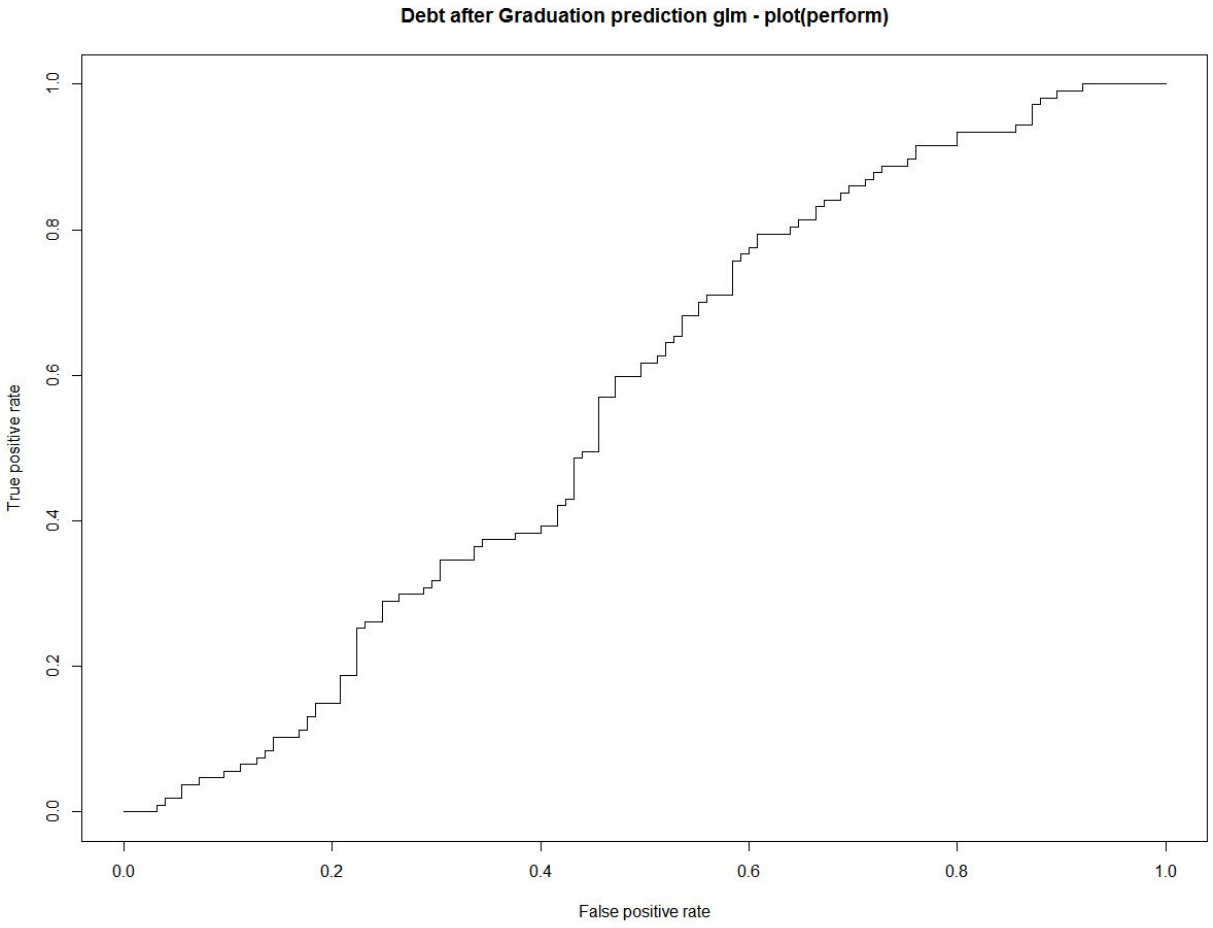
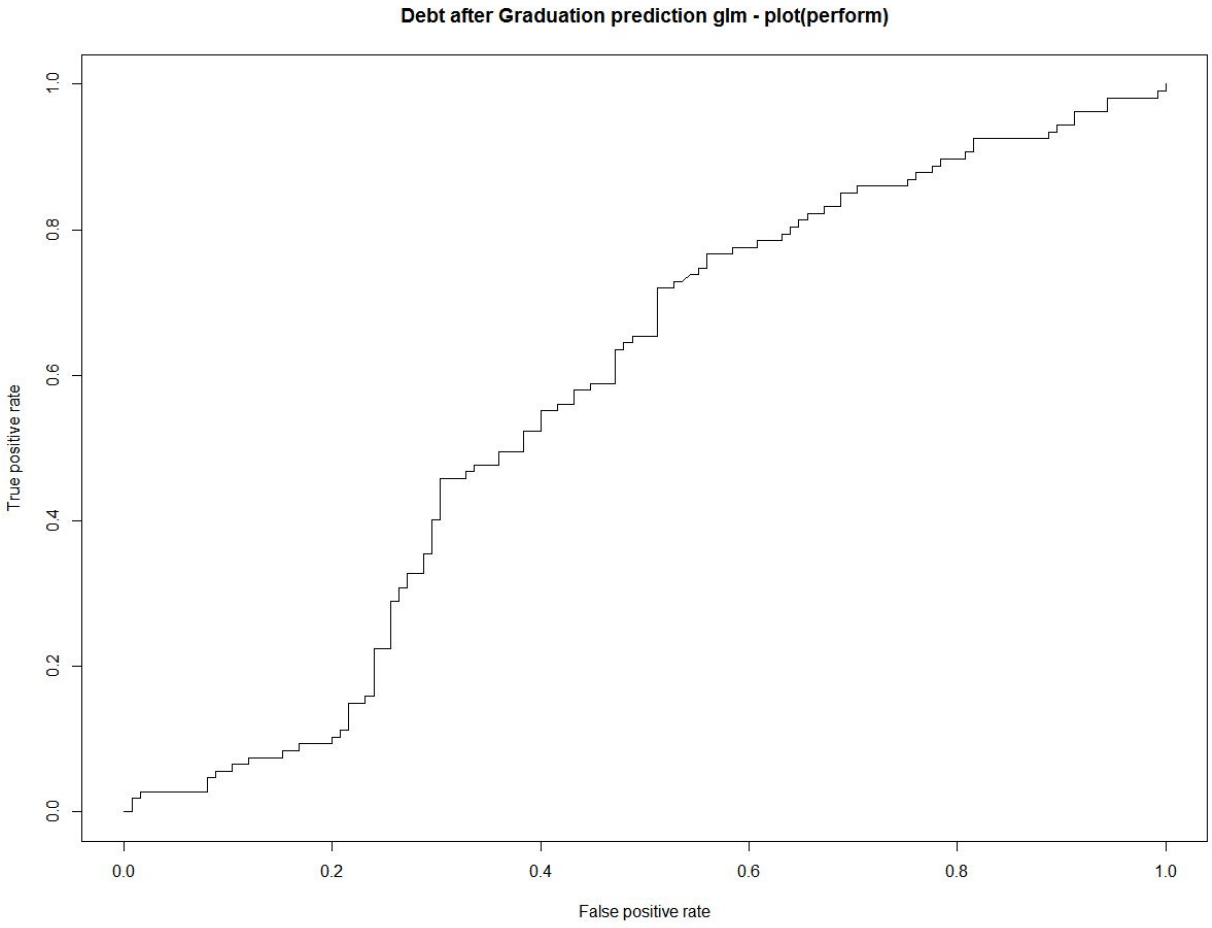
Dataset: https://www.kaggle.com/kaggle/college-scorecard
Data Preparation & Analysis Scripts are included in the zip file.
How-To-Help.txt is included in the zip file.
References.txt is included in the zip file.
Analysis & Summary-of-Findings are done below.
Presentation (PowerPoint) is included in the zip file.

Analysis was done by trying to classify whether or not a student would have greater than 25k in debt after graduation based on tuition, admission rate, and SAT midpoint scores of various colleges.

Plotting the performance of the AUC curve for the formula debtgt25k~ADM_RATE_ALL. The curve indicates that admission rate has no indication of what the debt upon graduation will be because the true positive rate has an almost 1:1 ratio with the false positive rate. So predictions will not be trustworthy on future data with this model.



Plotting the performance of the AUC Curve for the formula debtgt25k~TUITFTE. The curve indicates that tuition rate has no indication of what the debt upon graduation will be because the true positive rate has an almost 1:1 ratio with the false positive rate. So predictions cannot be trusted.
I find this odd because I had an initial assumption that higher tuition would lead to higher debt upon graduation. I will try another algorithm to see this linear model is not powerful enough.



Importance Plot of the fields for the random forest model. In both cases, the tuition is shown to have the greatest impact on accuracy and Gini index. The accuracy vs. Gini plot is surprising mainly for ADM_RATE_ALL. Although it does help with the accuracy of the model, it has a very large impact on the Gini index. In this case, the ADM_RATE_ALL having a high value for the Gini plot means that it helps with the branching of the decision trees to get to a leaf node where a yes/no classification can be made about whether a student's debt will be greater than 25k or not.