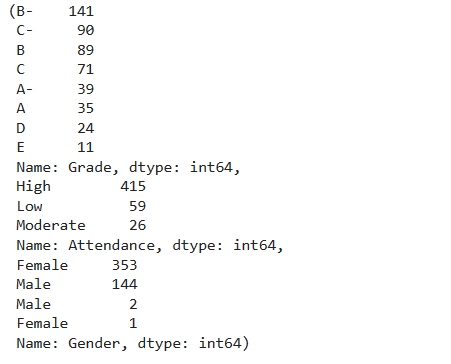


MTL 782 Assignment-1

1. **Data Composition:**



**2. Comparing Decision Tree accuracy with/without preprocessing categorical data**

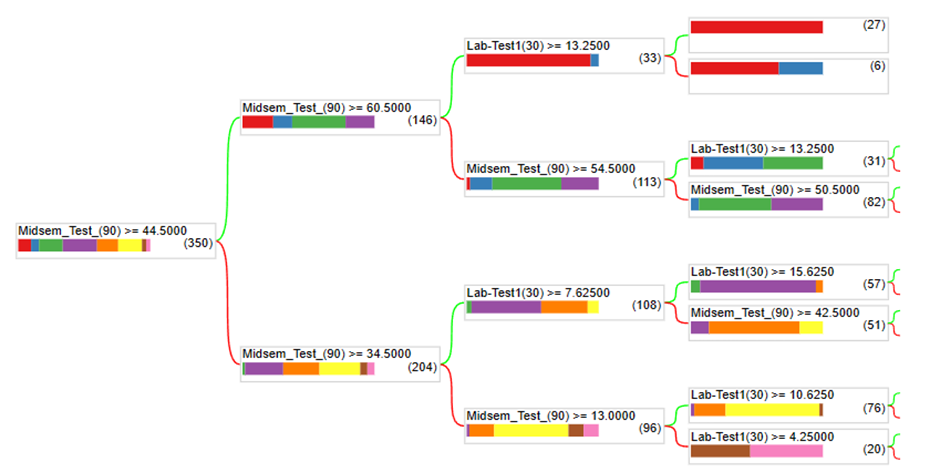
We notice that with preprocessing we are getting accuracy 0.80 and without preprocessing we are getting accuracy 0.791. We conclude that there is only a marginal error of 0.8% in accuracy.

Thus Tensorflow decision forests natively handle numerical and categorical.

**3. Random Forest Model on train-test split**

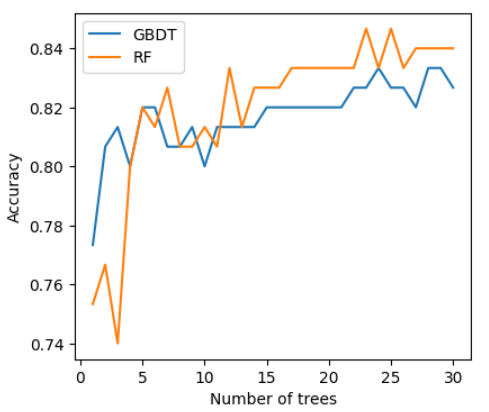
We obtained an accuracy of 85% with random-forest model involving train-test split of 70:30.

Visualization of the first tree in the trained dataset:

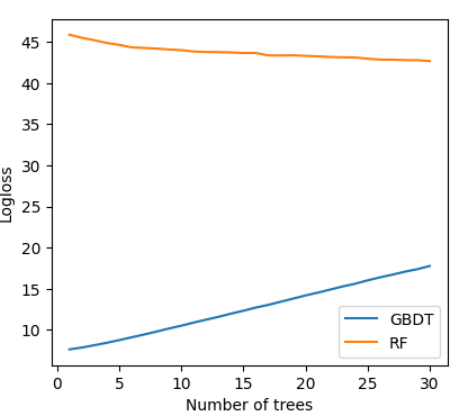
Iji

**4. Implementation of Gradient-Boosting Decision Trees & comparison with Random Forests**

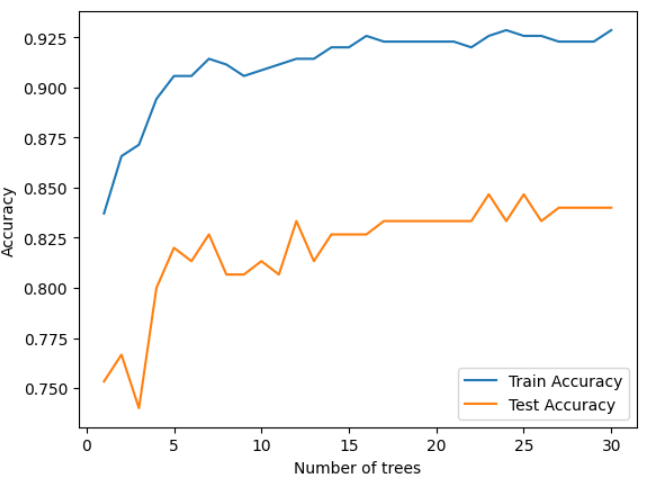
Accuracy curve as a function of no of trees:



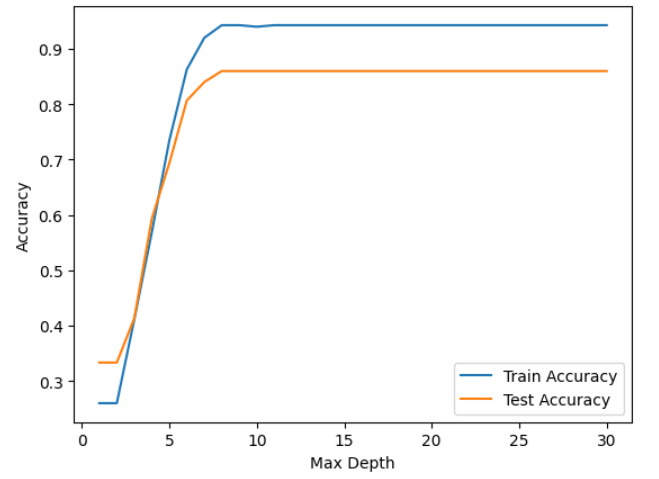
Log-loss curve as a function of no of trees:



### 5. **Train vs Test accuracies for Random-Forest with max\_depth and num\_trees hyperparameters**



As we can see increasing the no of trees parameter increases the accuracy of random forest model , the same is true in case of Max-depth hyperparameter as shown in the graph below:



Best Hyperparameters for accuracy of 85% or above using keras-tuner library:

Best hyperparameters: {'max\_depth': 29, 'num\_trees': 25}