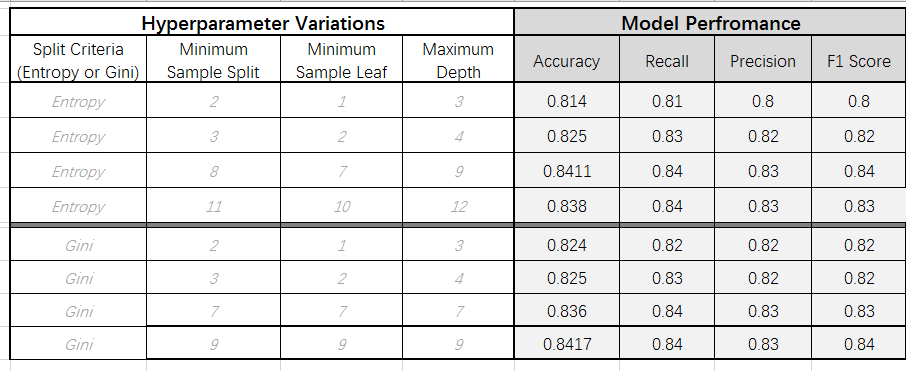
Q.1.1 Why does it makes sense to discretize columns for this prediction problem?

Decision tree model can deal with the discrete data.

Q.1.2 What might be the issues (if any) if we DID NOT discretize the columns.

It will increase the running time. Ordinary computers cannot handle such a large amount of computation. It will led a inaccurate results, such as overfiting.

Q.7.1 Decision Tree Hyper-parameter variation vs. performance (run your program  
manually for the following eight cases and enter the Model Performance values  
manually in the table)



Q.8.1 How long was your total run time to train the model?

About 79s.

Q.8.2 Did you find the BEST TREE?

Yes. Accuracy is 0.8417001064613873

Q.8.3 Draw the Graph of the BEST TREE Using GraphViz

  
Q.8.4 What makes it the best tree?

75% training and 25% testing. criterion='gini', min\_samples\_split=9, min\_samples\_leaf=9, max\_depth=9

Q.10.1 What is the probability of the outcome of the prediction for this ? What is  
your decision probability threshold and what is your predicted decision based on that?

The probability of less than 50k is 0.62, the probality of greater than 50k is 0.37

my prediction is less than 50k

Q. 10.2 What is the probability that your outcome prediction is accurate?

Use precision. The probability is 0.83