**Mandy He**

**CA03 – Decision Tree**

**Q1.1**

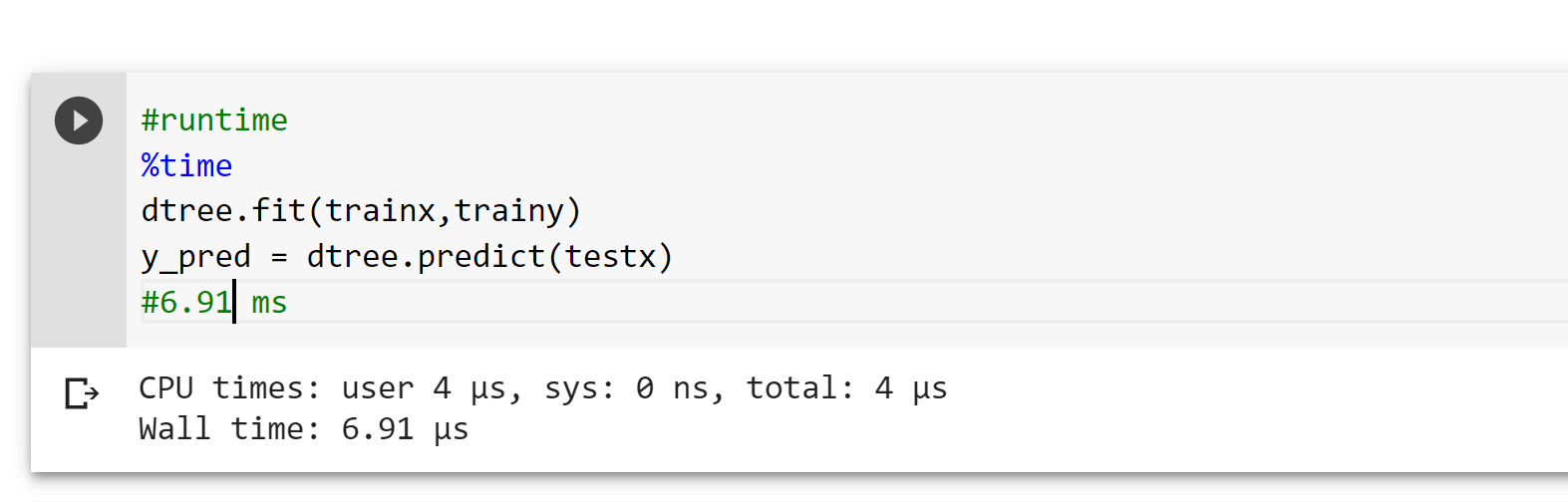
After completed the entire assignment, I believe that discretizing columns for the dataset totally make sense. Discretization is the process through which we can transform continuous variables, models or functions into a discrete form. Build decision tree algorithms model is a complicated process, using discretization can make our modeling clearer and easier to be explored.

**Q1.2**

The changing between each record would be so small if we didn’t do discretization. The training process will be slower and less effective if we keep using continuous data rather than using discrete data. Also, continuous data may have a skewed or non-standard distribution result.

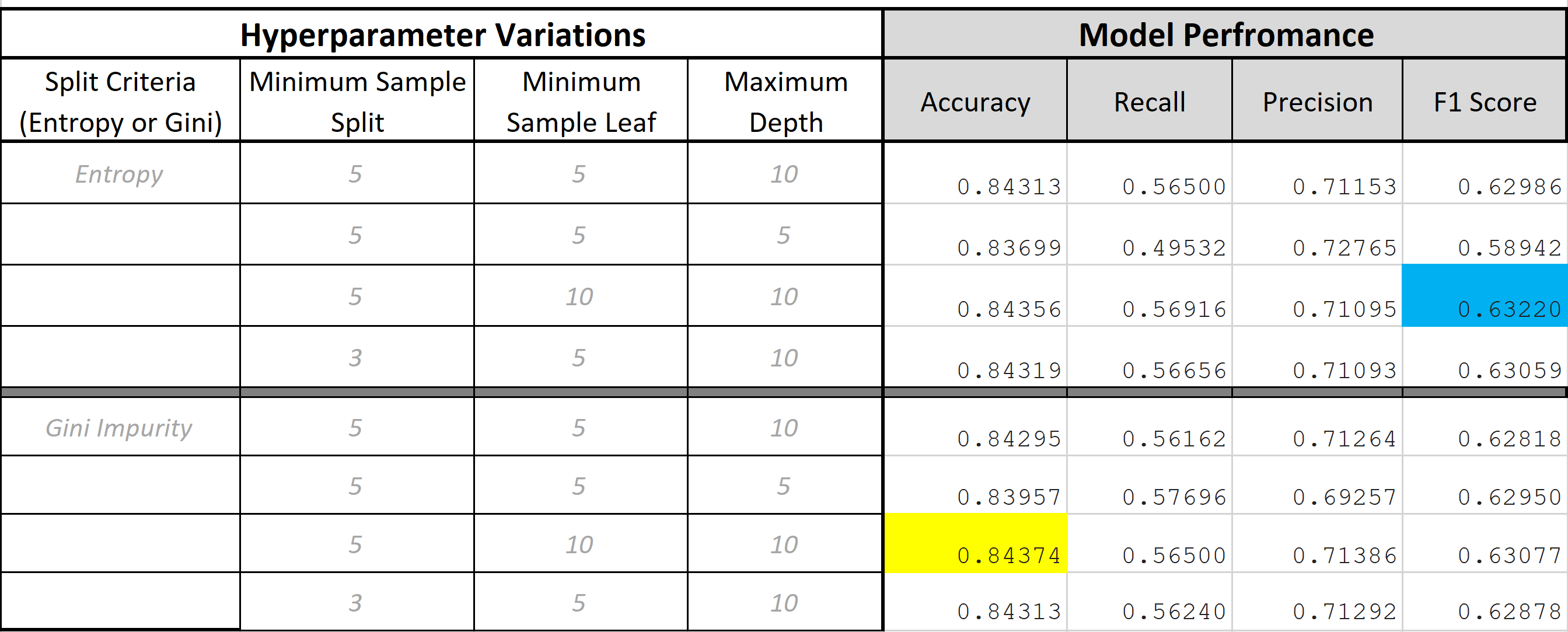
**Q8.1**

Running time is 6.91 ms

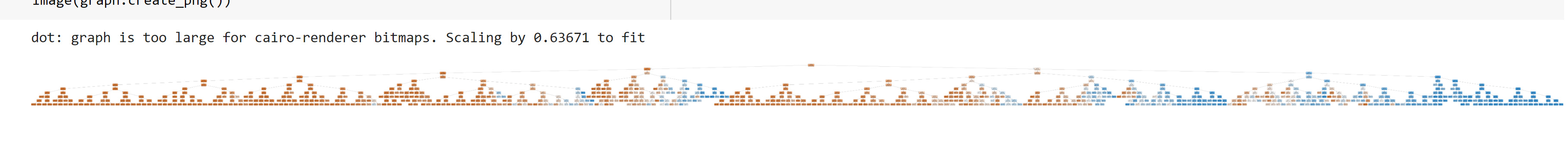


**Q8.2**

Based on hyperparameter variations model performance, we can tell that best model is Gini impurity with 5 as minimum sample split, 10 as minimum sample leaf, and 10 as maximum depth.



**Q8.3**



**Q8.4**

Since it has the highest accuracy score overall and F1 score is ok, which mean it performs good in terms of these 8 different tests.

**Q10.1**

The probability is 0.8435599778883361