Dhiman Sarkar
February 26th 2021
Introduction to Machine Learning
Dr. Arin Brahma
CA03

Q.8.1How long was your total run time to train the model?
My total run time to train the gini decision tree model took Wall time: 8.11 µs as per the below mentioned constraints:
max_depth=80,
random_state=101,
max_features= None,
min_samples_leaf= 2,
min_samples_split=21,
criterion=criteria

Q.8.2Did you find the BEST TREE?
Yes, I found the best tree with regards to Gini Impurity of Minimum Sample Split 21, Minimum Sample Leaf 2, Maximum Depth 80 giving the output of the highest F1-score 0.661 along with 0.857,0.591 and 0.75 as the respective Accuracy, Recall and Precision Scores

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

Q.8.4What makes it the best tree
A high f1 score predicts both positive and negative classifications. It has a good Harmonic Mean. On the contrary, precision scores opt for all positives including false positives. Whereas, recall scores opt for all negatives, including the false negatives. So invariably the false positives and false negatives get accounted. Therefore, a high f1 score will reduce the likelihood of that. Hence, the output with the highest f1-score should be a good indicator to identify the best Decision Tree.

Q.1.1 Why does it makes sense to discretize columns for this problem?
If we don't discretize columns for this problem, we will have too many groups if we didn't. This would create roadblocks for the decision tree to efficiently work.

Q.1.2 What might be the issues (if any) if we DID NOT discretize the columns.
We will have too many groups which is a major issue with respect to our classification model. Hence, we discretize the columns for the decision tree classifier model to work.