Luke Scott

COSC 311

Project Report

Dr. Wang

**Project Report**

Task 1: For parameters for the KNN model, I experimented with the n\_neighbors parameter, and found that using the value 4 provided the highest score for the model. For the DT model, I set the criterion parameter to entropy, then conducted a standard self-test on the data. The DT model scored perfectly compared to the KNN model that scored 99%, this is because the data wasn’t split.

KNN MODEL:

Calendar

Description automatically generated

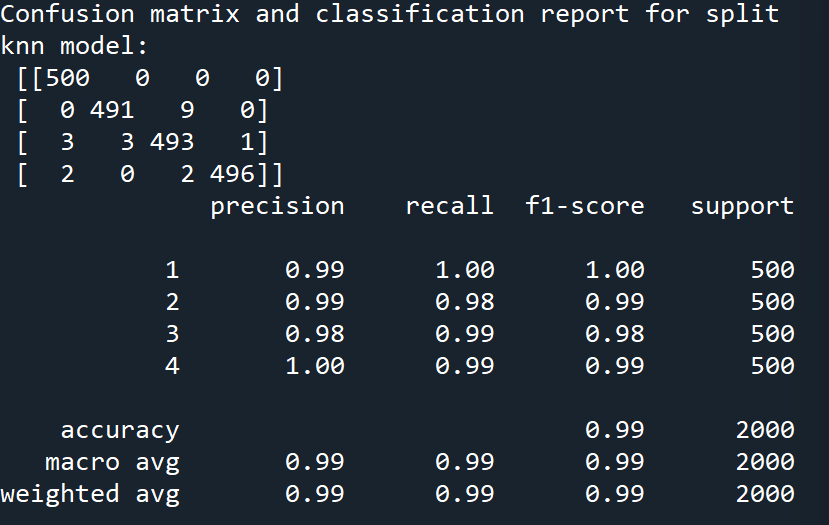
DT MODEL:

Calendar

Description automatically generated

Task 2: After the split, I kept the KNN parameters the same, however this model performed slightly worse compared to the KNN model performed before the split. I changed the parameters on the DT model as I wanted to change the max\_depth to 4 in order to give a proper number of classifiers for the data. The model performed worse when going above or below 4 for this value.

KNN SPLIT MODEL:



Chart, waterfall chart

Description automatically generated

DT SPLIT MODEL:

Calendar

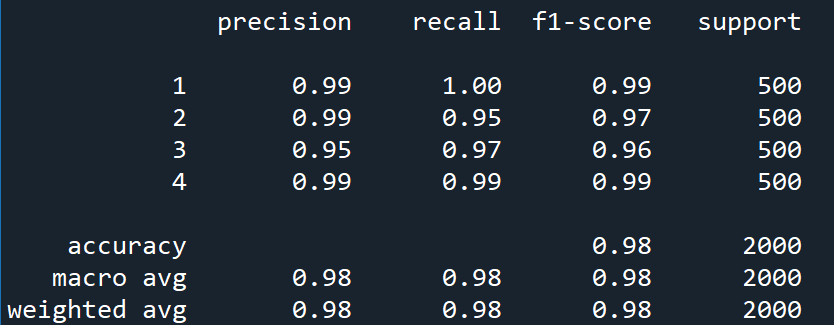
Description automatically generated

Chart, waterfall chart

Description automatically generated

Task 3: I chose to use the DT model to try to achieve the best results. Most modifications to the parameters didn’t improve the score, however I was able to slightly improve the score by increasing the max\_depth, changing the criterion to ‘gini’, and adding a min\_samples\_leaf and min\_samples\_split value of 5. This increased the score to almost 97%, which is still not better than the KNN score.

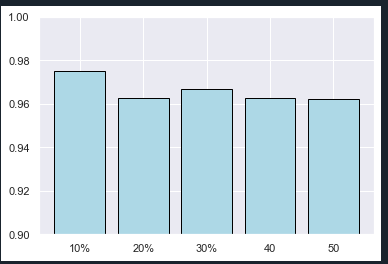
DT MOD:



Chart

Description automatically generated

The classification reports and confusion matrices for all the split tests (10%, 20%...) are shown in the .py file. Here is the bar figure displaying the difference in scores for these tests.



These test results are quite interesting because they decrease as the testing data pool is increased. With this data set, it seems that the model performs best with a smaller set of training data.