Mfolozi Dlamini

BSAN-6070 Intro to Machine Learning

2/20/2022

CA03 – Decision Tree Algorithm

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

**Binning is a technique for reducing the cardinality of continuous and discrete data. Binning groups related values together in bins to reduce the number of distinct values.**

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

**We would run into the issue of cardinality. We would end up with many distinct data points and would not be able to accurately make a prediction. Binning helps with the discretization of the data.**

QQ.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)

Table

Description automatically generated

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

**27 seconds on Google Colab**

Q.8.2 Did you find the BEST TREE?

**Yes.**

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



Q.8.4 What makes it the best tree?

**The best tree had an accuracy of 0.8432 (highest accuracy)**

**It has Recall, Precision & F1\_Score weighted averages of 0.84 (highest of all performance tuning efforts)**

Q.10.1 What is the probability of the outcome of the prediction for this?

**The probability of the prediction is [[0.91780822 0.08219178]]**

What is your decision probability threshold and what is your predicted decision based on that?

**Decision Probability threshold is 0.5 so the predicted decision based on that is [0]**

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

**The probability that the outcome is accurate is 0.8432 the same as the accuracy for our best model used to predict.**