Developing Classification Model for Diabetes from Health Indicators

Lillian Mueller lmuelle1@umd.edu

Abstract—

I. Introduction

II. Methodology

Listing 1. Libraries used for this assignment.

III. Results

Fig. 1. Decision Tree Utilizing Gini Impurity Criterion

Description	Train Data Accuracy	Test Data Accuracy	r2 Score
count	1000.0	1000	1000
mean	1.0	0.945100	0.915793
std	0.0	0.028636	0.044914
min	1.0	0.840000	0.750000
25%	1.0	0.920000	0.887892
50%	1.0	0.940000	0.916574
75%	1.0	0.960000	0.942824
max	1.0	1.000000	1.000000

TABLE I
STATISTICS FOR DECISION TREE UTILIZING GINI IMPURITY CRITERION

IV. Discussion

References

[1]

[2]