

Assignment 6

K nearest neighbors (KNN) to solve a classification problem

this is the data set I used in project 3 with some modifications to the data. In project 3 the predicted output (parental level of education) was simplified to if the parent did or did not attend college. In this test the parents education will remain categorical.

Data Set Source:

<https://www.kaggle.com/datasets/whenamancodes/students-performance-in-exams>

Parameters:

Input parameters math score reading score writing score	output parameter parental level of education Labels: some high school high school some college associate's degree bachelor's degree master's degree	Unused parameters gender race/ethnic group lunch test preparation course
--	---	--

preprocessing :

Applied standard scaler to input data

Results

K value = 3

Training

mean of metrics		precision	recall	f1-score	support
Accuracy: 0.98625					
sensitivity: 0.98625	associate's degree	0.96	1.00	0.98	164
specificity: 0.98625	bachelor's degree	0.97	1.00	0.98	91
F1 score: 0.9862667425995134	high school	1.00	1.00	1.00	160
log loss: 0.01971560764510944	master's degree	1.00	0.96	0.98	56
	some college	0.99	0.97	0.98	173
	some high school	1.00	0.98	0.99	156
	accuracy			0.99	800
	macro avg	0.99	0.99	0.99	800
	weighted avg	0.99	0.99	0.99	800

Testing

mean of metrics	precision recall f1-score support				
Accuracy: 0.18					
sensitivity: 0.18	associate's degree	0.18	0.18	0.18	39
specificity: 0.18	bachelor's degree	0.11	0.10	0.10	21
F1 score: 0.17843290810300025	high school	0.20	0.19	0.19	42
log loss: 19.287438161953293	master's degree	0.00	0.00	0.00	14
	some college	0.25	0.27	0.25	49
	some high school	0.17	0.17	0.17	35
	accuracy			0.18	200
	macro avg	0.15	0.15	0.15	200
	weighted avg	0.18	0.18	0.18	200

K value = 11

Training

mean of metrics Accuracy: 0.98625 sensitivity: 0.98625 specificity: 0.98625 F1 score: 0.9862667425995134 log loss: 0.01971560764510944	precision recall f1-score support
	associate's degree 0.96 1.00 0.98 164
	bachelor's degree 0.97 1.00 0.98 91
	high school 1.00 1.00 1.00 160
	master's degree 1.00 0.96 0.98 56
	some college 0.99 0.97 0.98 173
	some high school 1.00 0.98 0.99 156
	accuracy 0.99 800
	macro avg 0.99 0.99 0.99 800
	weighted avg 0.99 0.99 0.99 800

Testing

mean of metrics Accuracy: 0.18 sensitivity: 0.18 specificity: 0.18 F1 score: 0.17517587078880822 log loss: 4.694340804477007	precision recall f1-score support
	associate's degree 0.19 0.21 0.20 39
	bachelor's degree 0.17 0.14 0.15 21
	high school 0.18 0.21 0.20 42
	master's degree 0.00 0.00 0.00 14
	some college 0.20 0.20 0.20 49
	some high school 0.18 0.17 0.17 35
	accuracy 0.18 200
	macro avg 0.15 0.16 0.15 200
	weighted avg 0.17 0.18 0.18 200

K value = 100

Training

Accuracy: 0.98625 sensitivity: 0.98625 specificity: 0.98625 F1 score: 0.9862667425995134 log loss: 0.01971560764510944	precision recall f1-score support				
	associate's degree	0.96	1.00	0.98	164
	bachelor's degree	0.97	1.00	0.98	91
	high school	1.00	1.00	1.00	160
	master's degree	1.00	0.96	0.98	56
	some college	0.99	0.97	0.98	173
	some high school	1.00	0.98	0.99	156
	accuracy			0.99	800
	macro avg	0.99	0.99	0.99	800
	weighted avg	0.99	0.99	0.99	800

Testing

mean of metrics Accuracy: 0.215 sensitivity: 0.215 specificity: 0.215 F1 score: 0.20243645309673067 log loss: 2.539108198591018	precision recall f1-score support				
	associate's degree	0.16	0.21	0.18	39
	bachelor's degree	0.25	0.14	0.18	21
	high school	0.22	0.24	0.23	42
	master's degree	0.00	0.00	0.00	14
	some college	0.22	0.16	0.19	49
	some high school	0.25	0.40	0.31	35
	accuracy			0.21	200
	macro avg	0.18	0.19	0.18	200
	weighted avg	0.20	0.21	0.20	200