Kristen Kaiser

CS 422

Project 3

dataset used: Students Performance in Exams
https://www.kaggle.com/datasets/whenamancodes/students-performance-in-exams
data provided:

gender : { male || female}
race/ethnicity : {group A || group B || group C || group D || group E }
parental level of education: { associates degree || bachelor's degree || master's degree || some college || high school || some high school}
lunch: {standard || free\reduced}
test preparation course : { none || completed }
math score : int(range[0-100])
reading score : int(range[0-100])
writing score: int(range[0-100])

data pre-processing

The following modifications have been made to the data to test the hypothesis that a parent's college attendance can be predicted by a child's performance on their exams .

removed data for gender, race/ethnicity, lunch, and test preparation course.
Changed parental level of education to Boolean describing is parent college graduate.
1(TRUE) == { associates degree || bachelor's degree || master's degree}
0(FALSE) == {some college || high school || some high school}
Math score, reading score, and writing score.

Results via logistic regression

accuracy (test set): 0.705
sensitivity(test set): 0.34210526315789475
specificity(test set): 0.9274193548387096
f1 (test set): 0.46846846846846846
log loss (test set): 10.18897501838301
score (test set): 0.705

accuracy (train set): 0.6325
sensitivity (train set): 0.255663430420712
specificity (train set): 0.869653767820774
f1 (train set): 0.34955752212389374
log loss (train set): 12.693064292924092

score (train set): 0.6325

w = [-2.23047391] + ([0.01083734] * math score + [-0.04973253] * reading score + [0.06557917] * writing score)

Results via naïve Bayes

accuracy (test set): 0.69
sensitivity(test set): 0.5131578947368421
specificity(test set): 0.7983870967741935
f1 (test set): 0.5571428571428572
log loss (test set): 10.707120632101088
score (test set): 0.69

accuracy (train set): 0.59125
sensitivity (train set): 0.44660194174757284
specificity (train set): 0.6822810590631364
f1 (train set): 0.4577114427860697
log loss (train set): 14.11788077291863
score (train set): 0.59125