Yoseph Kebede

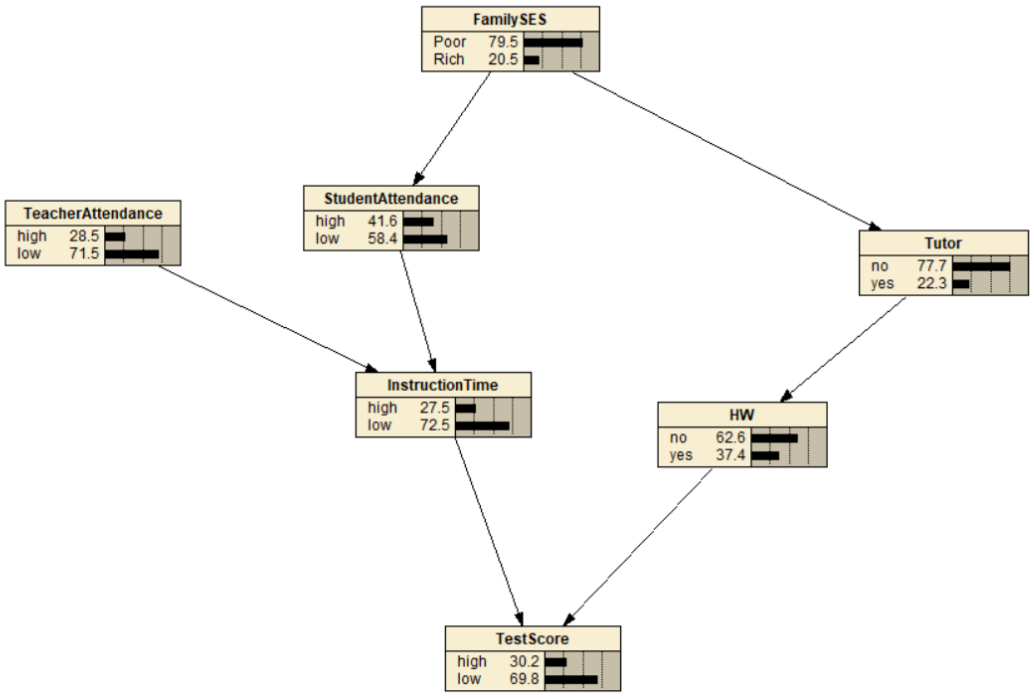
ENPM808Y

HW-9

Observation vs Intervention

Prompt: A simple narrative for the above model: Attendance of students from well off family is higher than the ones from poorer families. The chance of having a tutor for students from richer families is high. Students who have a tutor are more likely to do the homework. Student attendance and teacher attendance both affect time on instruction. Time on instruction and doing homework both affect the test scores.

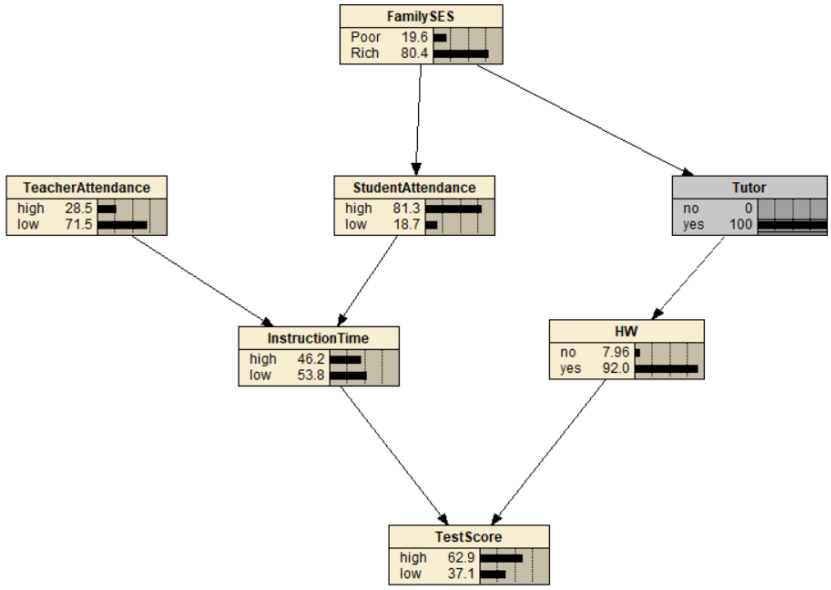
a. **Create this model in Netica and use the data of 500 students provided in this Excel file to learn the parameters (i.e., quantify the model)**.



b. What is the chance of a student scoring "high" in the test given the data (that is, without any observation)?

>> As seen from the figure above which is a Naïve Bayes Model created without any observation, the chance of scoring high is 30.2%

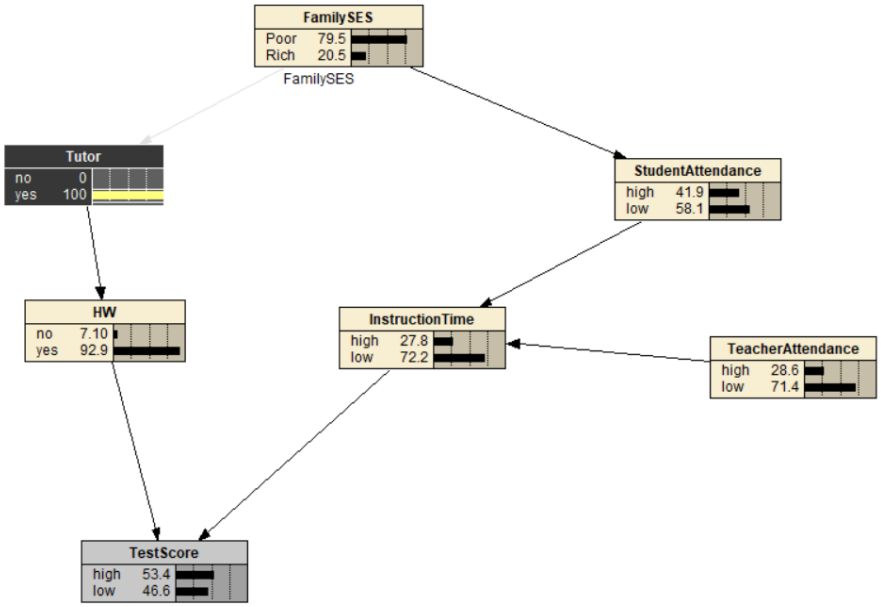
c. What is the chance of a student scoring "high" given that we observe that he has a tutor outside the class?



>> The chance of a student test score being high given the student has a tutor is 62.9%

d. Having what observed the above and thinking that having a tutor outside the class would improve learning, the principal decides on an intervention to provide tutors for all students. To estimate she would like to estimate the causal effect of her intervention. Given this intervention, what is the chance that a student scores high on the test and how is it different from (b), that is, when there was no intervention? (Remember to use Enter Action in Netica instead of Enter Finding)?

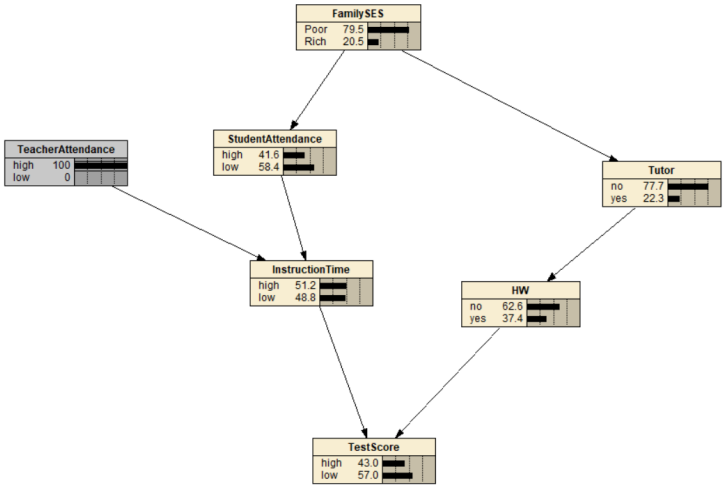
>> a. The student scores high on test given the intervention on tutor with a chance of 53.4%
When there is no intervention.



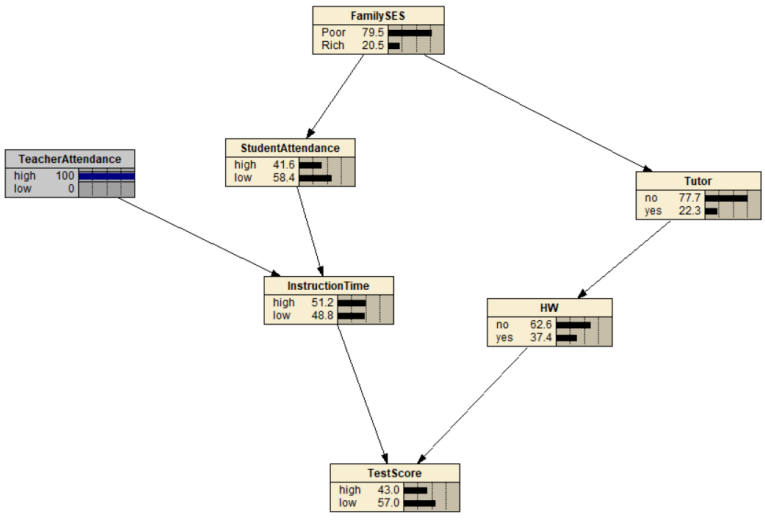
e. Why are the results in c and d different?
>> For part c the chance of a student having a tutor is a given whereas for part d every student has a tutor; thus the probability of a student being poor or rich has no factor on a tutor for part d. Therefore, since the family social economic status plays no role in determining the presence of tutor ergo homework production goes high. As a result, high and low test cores become about equal.

f. What can you say about a different intervention such as an intervention on teacher attendance as opposed to observing that teacher attendance is *high*?

>> given attendance high



intervention

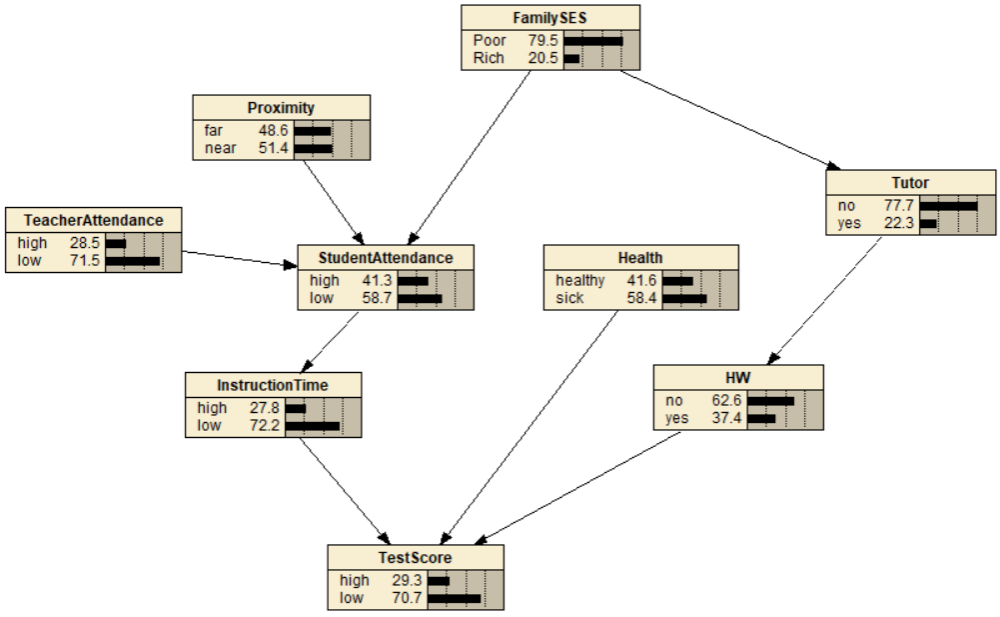


There is no difference between intervention or observation on teacher attendance since it's not linked with any node.

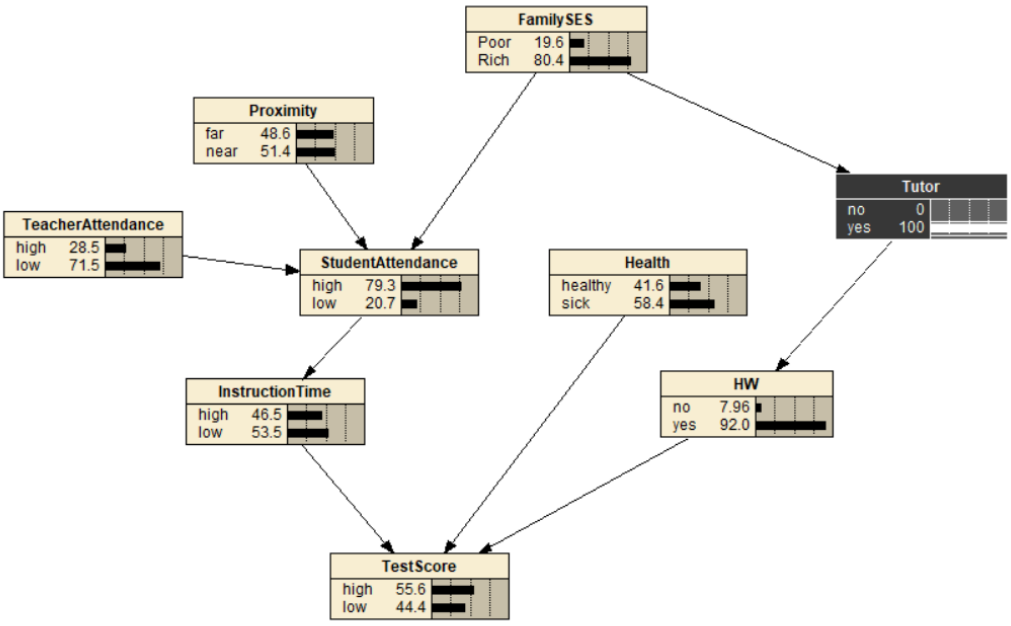
g. Now consider the fact that *student's health* and *home proximity* to school are also important factors. Extend the model shown above to include these two factors. Explain your reasoning.

>> Proximity has been randomized so that this model assumes that students regardless of their economic status live near or farther from their schools, ie disregarding the general notion that well to do schools exist near rich communities. Additionally, again disregarding the general intuition that rich students get sick fewer times than poor students, the health node has been added to the model to consider the students' health status on the day of the exam. Therefore, the below scenario considers the effect of these two nodes without any observation.

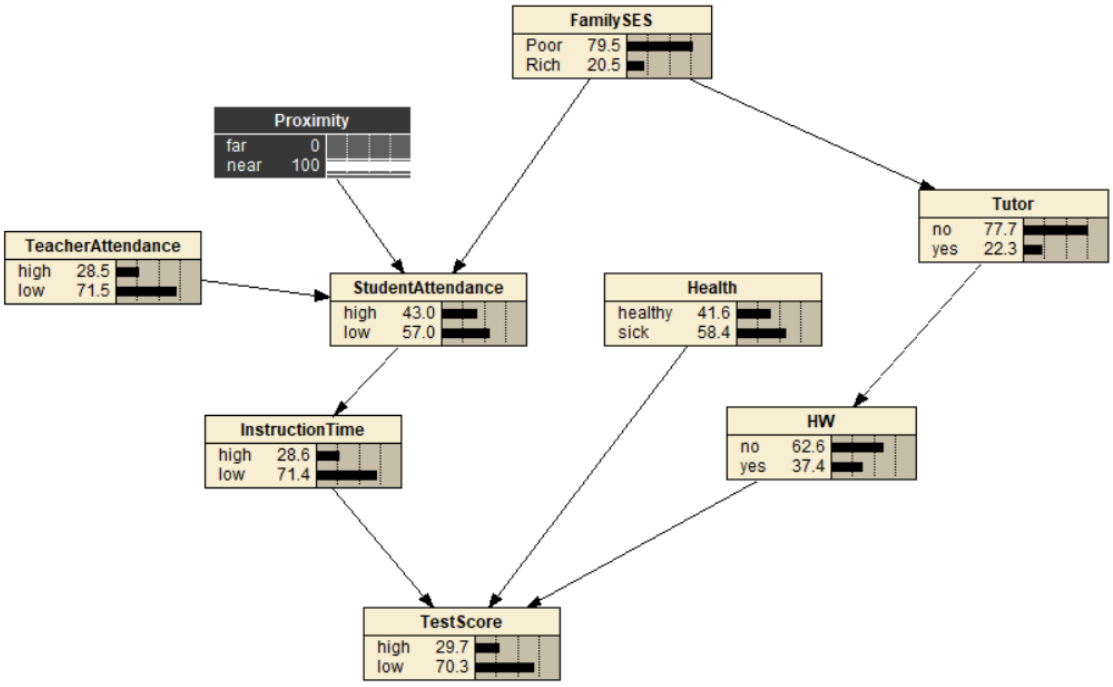
Thus, results show that without any observation or intervention, the chance of students scoring high has reduced to 29.3% when these variables were added (in this Naïve Bayes model), almost similar to the model we started with that lacks these to nodes (ie 30.2%). For comparison sake, tutor, health and proximity will be observed one at a time to see the difference in high test score outcomes while these two variables have been added.



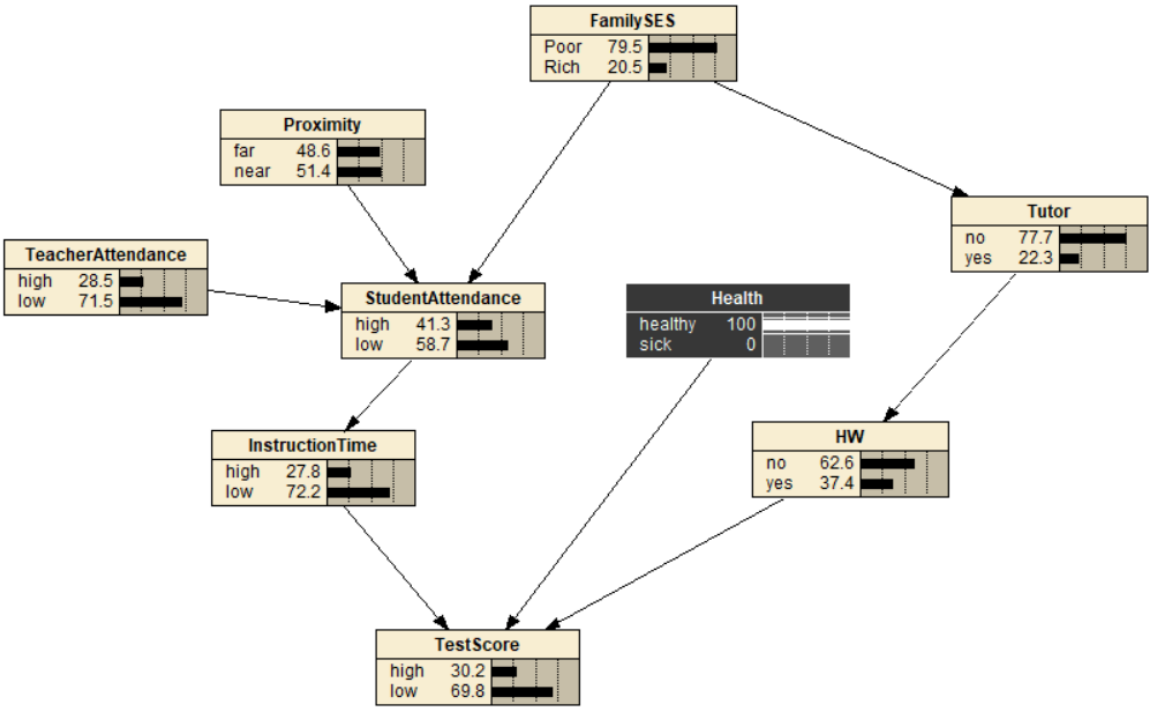
A. Tutor observed as yes. Figure below shows that high test scorers rose to 55.6%



B. Proximity given near. Brought no significant difference, only 29.7% up from baseline by 0.4%



C. Health given that all are healthy. Similarly the high scorers only bumped by 0.9% to 30.2%



Therefore, since in my update of health and proximity, the students who were listed on either side of the classifier were of even numbers, there was no significant difference seen on test scores. However, with the change in data to have data points skewing to one or the other side, in addition to combining the other nodes, a wide array of predictive results can be obtained.