1. Weight: Higher or the lower the weight of an individual is, the value of the BMI changes. Also, this variable does play a direct role In the BMI since it is a factor that is used for the BMI calculation.

Standing Height: Height of an individuals a second factor that contributes to the BMI value. For example, the taller the individual is the higher the value of their BMI. Standard height also plays a major role in the computation for the BMI.

Direct HDL-Cholesterol: Poor diet can lead to higher amount of Cholesterol in the blood of the individual and this can lead to higher weight, which contributes to BMI value change.

Number of months working in the main job: Choose this variable, because working in the work force may lead to higher stress level in the body of the individual that may contribute to higher chances of BMI value.

Age: Because it seems more likely that an order person might have a higher BMI.

Days physically active at least 60 mins: This is because it tends to the fact that people with some physical actives tend to have a normal BMI, not higher or lower. People with less amount of workout may gain more weight and HDL- Cholesterol, which are some key evidences of higher bmi.

A screenshot of a computer

Description automatically generated

Upon doing a couple of “Split Best”, it is noticeable that the AIC value is decreasing, and the r square value is increases. Although the changes are not significant. However, the more the number of split the better the decision tree gets, due to the higher R square value with a lower AIC value.

On the left side of the decision tree, the mean and the standard deviation are lower than the child nodes on the right. The mean and the standard deviation for the child nodes are double the value when compared to the child nodes on the left side. Also, the mean and the standard deviation between child nodes are unique values.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

I noticed that this decision is slightly worse than the previous one, because when looking at the AIC value and the R square, it is noticed that it is higher in the decision tree that has the variable “Waist Circumference”. Although, just by doing the “Split best” I see that the AIC value is decreasing and the R square value is increasing just like the previous model, which is a good thing.

For the mean and standard deviation for the child nodes, I would say that they are unique for each node. The values tend to be like the previous decision tree. However, the mean value is higher on both ends. Unlike the previous decision tree, where the left side had smaller value than the ones on the right.

I think it is because the decision tree with the X factor of “Waist Circumference” would be more complex than the one without, which makes the AIC value higher. One column might be loading on to the other increasing the complexity of the dataset. The other reason could be because the model is trying to overfit, by trying to cover irrelevant patterns in the dataset. Which makes it hard to use the model for prediction/validation, since the model is not ready. In the sense that it is not trained enough for validation. One difference I noticed between the decision trees of numeric target verse the categorical target, is that in the numeric target there is the AIC value, however in the categorical target there is not one. However, it is evident that the more number of predictors in the X factor the more complex the decision tree gets for both the categorical and numeric target.