* What is the variance in this scenario?

The variance is 4%. In this scenario, the variance would be the difference between the training score (79%) & the testing score (75%).

* What may happen to the bias error if variance increases?

As variance increases, the model pays too much attention to the training details, capturing noise as if it's a real pattern. This can cause overfitting, where the model fits the training data too closely. While high variance might make the model perform well on training data, it often leads to poor performance on new, unseen data.

* What is the use of validation data?

Validation data is crucial for evaluating a model's performance during training and adjusting its hyperparameters. It acts as a fresh set of data unseen during training, aiding in estimating how well the model can generalize to new, unseen data. Using validation data helps prevent overfitting and assists in choosing hyperparameters that optimize the model's performance on unseen data.

* After tweaking the hyperparameters, you observe that the training score increases to 81% and test score to 78%. Is this better than the previous model? Why?

Yes, the adjusted model with a training score of 81% and a test score of 78% is better than the previous model. The increase in both training and test scores indicates improved performance and generalisation. The smaller gap between the training and test scores (3% compared to the previous 4%) suggests that the model is likely better at generalizing to new data, reducing overfitting, and making it a more robust and reliable model for unseen instances.