

Here is the generated checklist:

Data Allocation Checklist

Data Set Separation

- Are three separate data sets created: training, validation, and test?
☐ Yes / No
- Is the test data set independent from the other two data sets?
☐ Yes / No
- Are the data sets complete and representative of the population?
☐ Yes / No

Model Family Selection

- Has bias-variance trade-off been considered in model family selection?
☐ Yes / No
- Is the complexity of the selected model family high enough to minimize bias, but not too high to avoid high variance?
☐ Yes / No
- Are methods used to provide estimates of the bias and variance of the selected model family (e.g. holdout data, k-fold sampling)?
☐ Yes / No

Learning Process Verification

- Has the trained model been evaluated using the test data set?
☐ Yes / No
- Is the performance of the trained model documented?
☐ Yes / No
- Have systematic errors in the training data set been identified and mitigated?
☐ Yes / No

I hope this helps! Let me know if you have any further requests.