Machine Learning Model Handbook Grading Rubric

Part 2: Data Handling

Exercise 2b

Describe your exploratory data analysis of any target and input features of note. Include the following:

- How many rain and snow records are in the dataset?
- Do the distributions of values make sense for the physical world?
- Are there any unexpected values?
- Which input features may be the strongest predictors of rain vs snow?
- Include any important plots to illustrate your conclusions. Limit yourself to 5 plots.

Score	Criteria
5 - Excellent	• Correctly identifies the number of rain and snow records • Evaluates if variables fall within expected ranges and identifies major outliers with plausible justification. • Identifies the strongest correlations between variables and precipitation type. • Supports analysis with relevant plots and explains their significance.
4 - Proficient	• Addresses most supporting questions with clear reasoning. • Identifies expected value ranges and major outliers, though some discussion may lack depth or explain deviations. • Discusses variable and precipitation type correlations, but may not fully justify conclusions. • Includes relevant plots but may not thoroughly explain all of them.
3 - Satisfactory	• Discusses expected value ranges but may overlook key outliers. • Addresses variable and precipitation type correlations but lacks

	strong supporting evidence. ● Includes some plots but does not clearly explain their significance.
2 - Needs Improvement	Little to no evaluation of expected ranges or outliers. ● Mentions variable and precipitation type correlations but with weak or no supporting analysis. ● Few or no relevant plots included.
1 - Minimal	● Does not adequately analyze expected value ranges, or outliers. ● No meaningful discussion of variable and precipitation type correlations. ● Little to no supporting plots.
0 - No Response	● No response or entirely off-topic answer.

Part 3: Model Development

Exercise 3e

Paste evaluation results

Then describe the results of the original model validation. Include the following:

- How well does the model predict rain? Support your description with the evaluation metrics.
- How well does the model predict snow? Support your description with the evaluation metrics.
- How do you interpret these statistics in the context of the physical world?
- What changes will you make to try to improve these statistics in the next iteration?

Score	**Criteria**
5 - Excellent	● Includes evaluation results. ● Provides a well-reasoned interpretation of validation metrics (accuracy, precision, and recall) for rain and snow. ● Connects evaluation results to real-world meteorological implications, such as the implications of incorrect classifications in weather prediction. ● Suggests potential improvements for subsequent trials with supported reasoning.

4 - Proficient	● Includes evaluation results. ● Evaluates model performance for rain and snow using accuracy, precision, and recall with mostly correct interpretations. ● Connects evaluation results to real-world meteorological situations, but may lack sufficient descriptions of real-world implications. ● Suggests reasonable improvements, though some may lack depth.
3 - Satisfactory	● Includes evaluation results. ● Provides a basic interpretation of validation metrics but may miss some key details or misinterpret one or more metrics. ● Mentions real-world situations but does not fully explore the impacts of incorrect classifications. ● Suggests general improvements but lacks clear justification.
2 - Needs Improvement	● Includes evaluation results. ● Attempts to interpret validation metrics for rain and snow but contains inaccuracies or lacks depth. ● Connections to real-world situations are weak or missing. ● Suggestions for improvement are vague or missing.
1 - Minimal	● Includes evaluation results. ● Provides an incorrect evaluation of model performance with little to no reference to accuracy, precision, or recall. ● Connections to real-world situations are weak or missing. ● No actionable suggestions for improvement.
0 - No Response	● No response or entirely off-topic answer.

Exercise 3f

Paste the full output of each of your validation trials, one per box.

Score	**Criteria**
5 - Excellent	● At least three unique additional trials are present ● Trials include more than one algorithm ● Trials use a variety input features
3 - Satisfactory	● Fewer than three unique additional trials are present
0 - No Response	● No response or entirely off-topic answer.

Exercise 3h

Then make a final decision on whether this model delivers on the results needed with supporting justification. Include the following:

- Which precipitation class(es) had the best evaluation metrics? List some physical scientific reasons why this may be the case.
- Is this model ready for use in the real world? Why or Why not?
- What other possible changes could further improve this model?

Score	**Criteria**
5 - Excellent	● Accurately identifies the precipitation class(es) that had the best evaluation metrics given their choices in model development. ● Suggests reasonable physical scientific reasons why these variables generated the best model performance, including any surprising results. ● Thoughtfully assesses whether the model is ready for real-world use, providing strong justification within the context of the scientific issue at hand and the initial problem statement. ● Suggests concrete, scientifically valid improvements for future iterations.
4 - Proficient	● Accurately identifies the precipitation class(es) that had the best evaluation metrics given their choices in model development. ● Suggests reasonable physical scientific reasons why these variables generated the best model performance, but may lack depth in some areas. ● Assesses real-world readiness with justification, though reasoning may not be fully developed. ● Suggests meaningful improvements, but they may not be fully explained.
3 - Satisfactory	● Identifies the precipitation class(es) that had the best evaluation metrics, but with partial accuracy or missing details. ● Offers a basic scientific explanation for model performance but lacks depth. ● Provides a general assessment of real-world readiness, though justification is weak or incomplete. ● Suggests potential improvements but with little scientific reasoning.

2 - Needs Improvement	● Attempts to identify the precipitation class(es) that had the best evaluation metrics, but with significant inaccuracies or missing key metrics. ● Provides little or unclear scientific reasoning behind model performance. ● Minimal discussion of real-world readiness, with weak or unsupported justification. ● Suggestions for improvement are vague or not scientifically valid.
1 - Minimal	● Fails to correctly identify the precipitation class(es) that had the best evaluation metrics. ● Fails to assess real-world applicability. ● Offers little to no discussion on real-world readiness. ● No meaningful suggestions for improvement.
0 - No Response	● No response or entirely off-topic answer.