03 Analysis - Supervised Machine Learning Readiness



Machine Learning Model Handbook

Name(s):

| Part 1: Problem FramingExercise 1Which scientific question should be answered? |
| --- |
|  |

| Part 2: Data HandlingExercise 2bDescribe your exploratory data analysis of any target and input features of note. Include the following:* Do variables follow diurnal or annual patterns generally as expected?
* Do the variables have the expected ranges of values? Do any variables appear to include major outliers?
* Which stations appear to be most correlated to the variables at Mt Mitchell?
* Include any important plots. Limit yourself to 5.
 |
| --- |
|   |

| Exercise 2cInput your data splitting strategy below. |
| --- |
| Training % |  |
| Validation % |  |
| Testing % |  |

| Part 3: Model DevelopmentExercise 3ePaste your evaluation metrics below. |
| --- |
|  |

| Describe the results of your initial model validation. Include the following:* Which variables have favorable evaluation metrics? Which variables don’t perform as well?
* 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?
 |
| --- |
|  |

| Exercise 3fPaste the full output of each of your validation trials, one per box. You may complete as many trials as you like until you are satisfied with the evaluation metrics, or they no longer improve with new trials. *If you need to add new boxes, right click the bottom-right box and select* ***Insert row below*** |
| --- |
|  |  |
|  |  |
|  |  |

| Exercise 3hPaste your testing evaluation metrics below. |
| --- |
|  |

| Describe how your testing metrics compare to your validation metrics. Include the following:* Which environmental variables 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?
 |
| --- |
|  |