Reading & Discussion: Chapter 5 - Machine Learning (Validation Pages 359-375)

TASK

Your task for this discussion is to read [Chapter 5 - Machine Learning (Validation Pages 359-375)Links to an external site.](https://jakevdp.github.io/PythonDataScienceHandbook/05.03-hyperparameters-and-model-validation.html). This can be found in the link provided or the [pdf version of the textbook](https://elearning.mines.edu/courses/52392/files/5459013/download?wrap=1)[Download pdf version of the textbook](https://elearning.mines.edu/courses/52392/files/5459013/download?download_frd=1)for the class that we have provided. You will then post a discussion and a reply below!

DISCUSSION REQUIREMENTS

Once you have completed this weeks reading, you are asked to complete the following tasks:

1. Post a 1-2 sentence response from the reading for 2 of the three below prompts:

a. What happens when you train and validate your model on the same data? Why is this bad?

b. How do you split the data into training and testing sets, so you can properly validate your model?

c. Provide a free-response: this can be anything from clarification questions, something that piqued your interest, or maybe a personal experience you have with what was discussed in the book section.

2. Post a meaningful reply to another student's question/post.

Our goal is to help you get a deeper understanding of data science and discussions are a wonderful way to do this! Think of this as a learning opportunity for both you, your classmates, and your professor! Be sure to use academic language and question ideas from the readings for clarity or to push your own thinking.  You may of course respond to more than two of the questions and more of your classmates if you would like.

# **Instructional Videos & Learning Materials: Model Validation**

## WATCH

Watch this week's video, download the working files so you can follow along and experiment with the platform we'll be using for the course.

**Model Validation Video**

## WORKING FILES

You will need the following files to get set up for this week. They include the following:

* Jupyterhub notebook (download to a personal device then upload to your Jupyterhub). Also, download the image and place it in the same directory in your Jupyterhub notebook.
  + [Model Validation Jupyter Notebook](https://elearning.mines.edu/courses/52392/files/5459097/download?wrap=1)[Download Model Validation Jupyter Notebook](https://elearning.mines.edu/courses/52392/files/5459097/download?download_frd=1)
  + [Supporting Image](https://elearning.mines.edu/courses/52392/files/5459030/download?wrap=1)[Download Supporting Image](https://elearning.mines.edu/courses/52392/files/5459030/download?download_frd=1)
* PDF file that contains the same information as the Jupyterhub notebook, in case your Juypterhub is not yet working.
  + [Model Validation PDF](https://elearning.mines.edu/courses/52392/files/5458999/download?wrap=1)[Download Model Validation PDF](https://elearning.mines.edu/courses/52392/files/5458999/download?download_frd=1)

## ADDITIONAL RESOURCES

The links below are supplementary resources to help you review or strengthen the topics we have discussed:

* [Cross ValidationLinks to an external site.](https://www.pluralsight.com/guides/validating-machine-learning-models-scikit-learn)
* [Grid SearchLinks to an external site.](https://www.mygreatlearning.com/blog/gridsearchcv/)

# **Instructional Videos & Learning Materials: Model Validation**

## WATCH

Watch this week's video, download the working files so you can follow along and experiment with the platform we'll be using for the course.

**Model Validation Video**

Activity: Confusion Matrix

BACKGROUND

Today you will be working on completing a confusion matrix worksheet to solidify your understanding of confusion matrices. These matrices are important to be used in evaluating your classification model, sometimes using a single metric (like accuracy) can produce misleading results that could give you higher confidence in a bad model!

TASK

1. READ: this "[Why accuracy alone ... Links to an external site.](https://tryolabs.com/blog/2013/03/25/why-accuracy-alone-bad-measure-classification-tasks-and-what-we-can-do-about-it/)" article that provides an introduction with a good example of why we want to look at a holistic view of our results versus a single metric.

2. WATCH: the below introductory video about confusion matrices.

**Confusion Matrix Introduction Video**

. COMPLETE: your own confusion matrix worksheet. Congrats!! You now have a better understanding of how confusion matrices work and what their uses are for! Now it is time to transfer this new knowledge to paper so you can remember it!

**Print** out the [PDF version of the worksheet](https://elearning.mines.edu/courses/52392/files/5458992/download?wrap=1)[Download PDF version of the worksheet](https://elearning.mines.edu/courses/52392/files/5458992/download?download_frd=1)and fill it in.

**Confusion Matrix Worksheet Video**