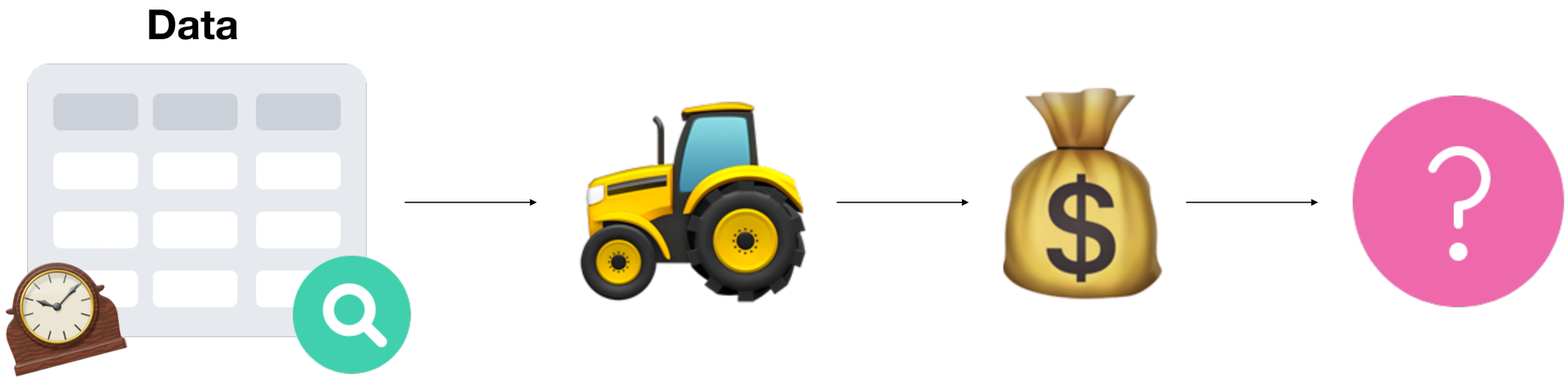
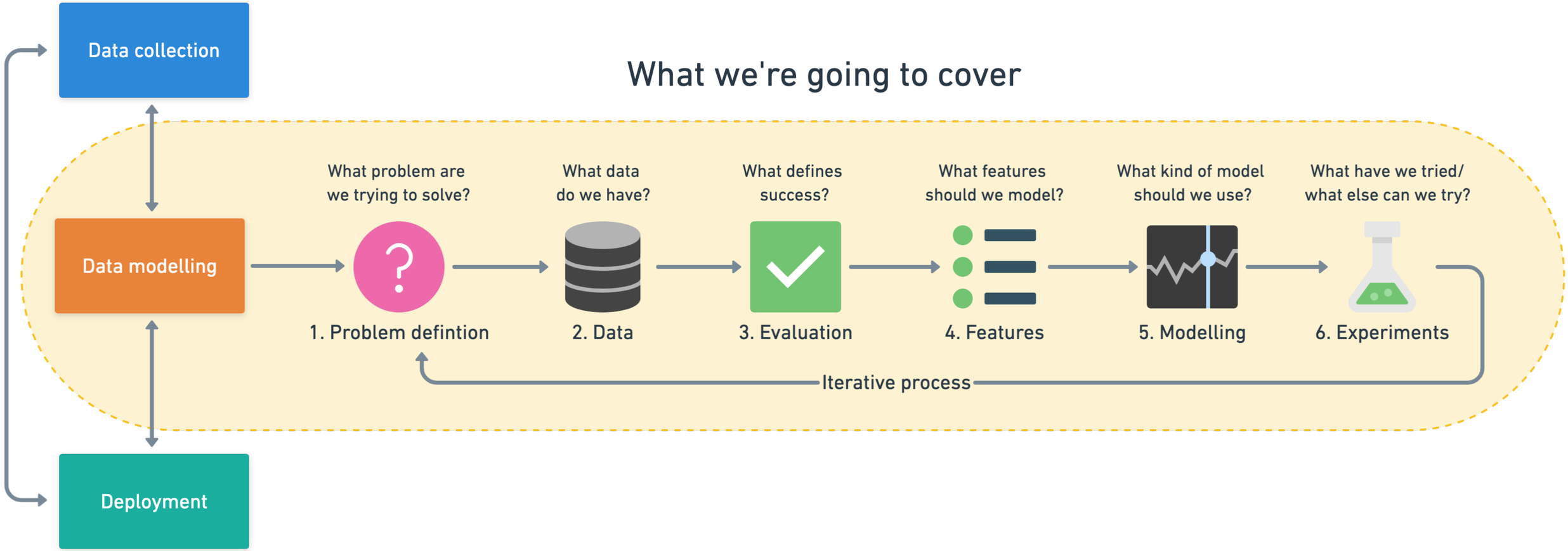
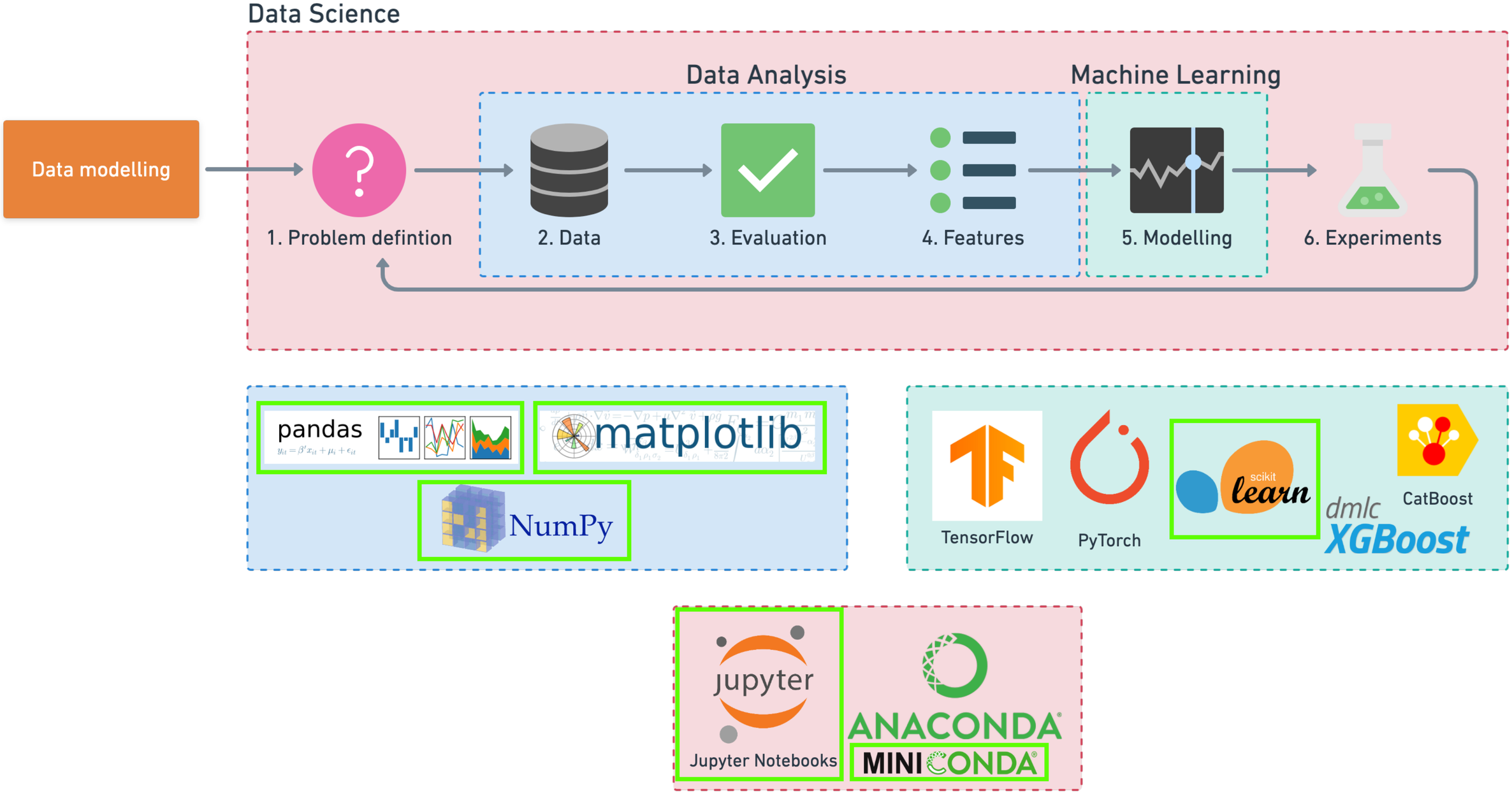
Structured Data Project 2: Predicting the sale price of Bulldozers (regression)



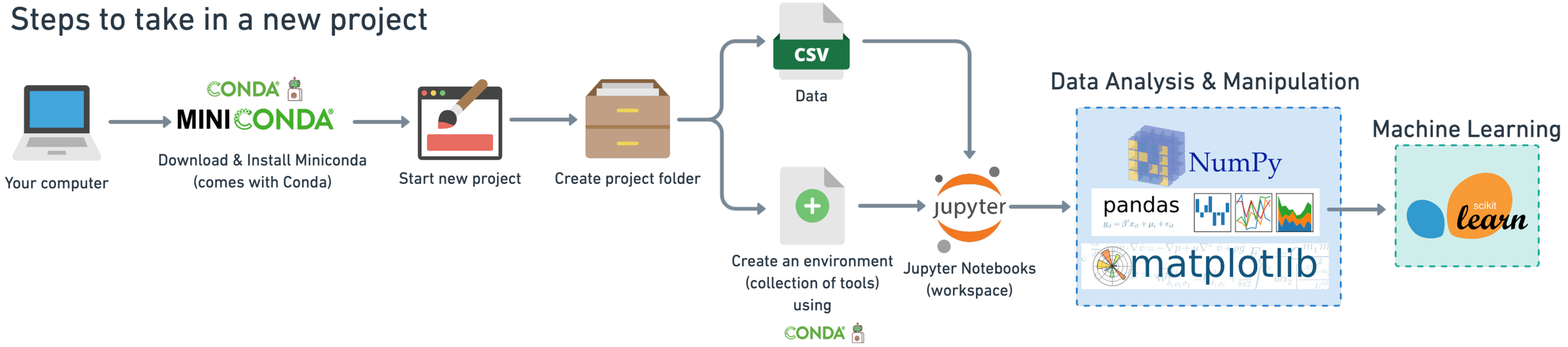
Steps in a full machine learning project



Tools you can use

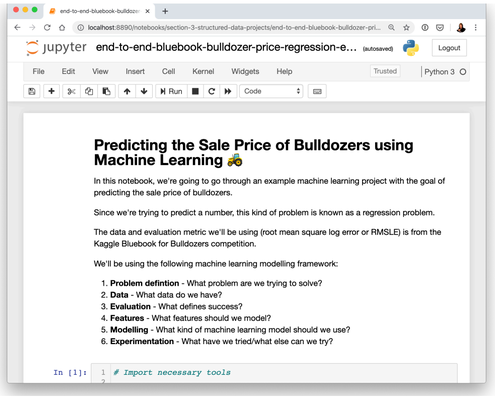


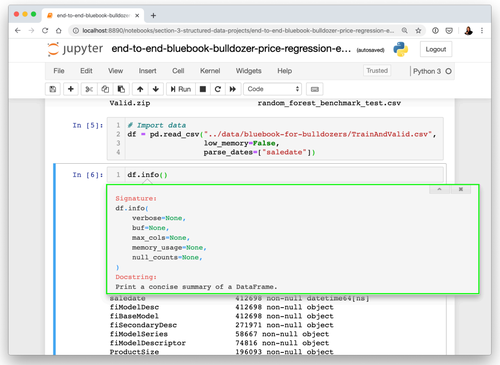
Steps to take in a new project

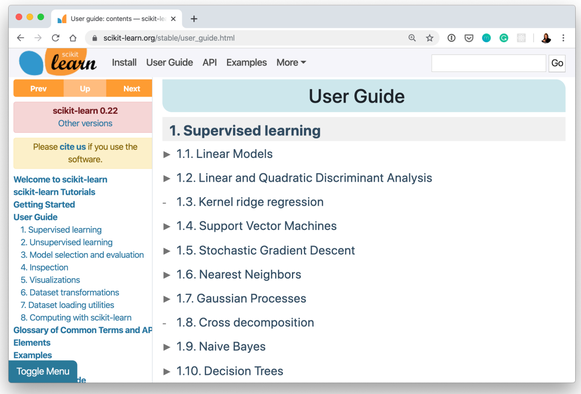


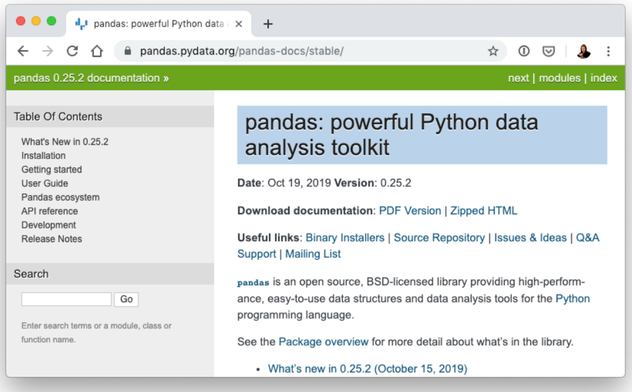
Where can you get help?

- Follow along with the code
- Try it for yourself
- Press SHIFT + TAB to read the docstring
- Search for it
- Try again
- Ask



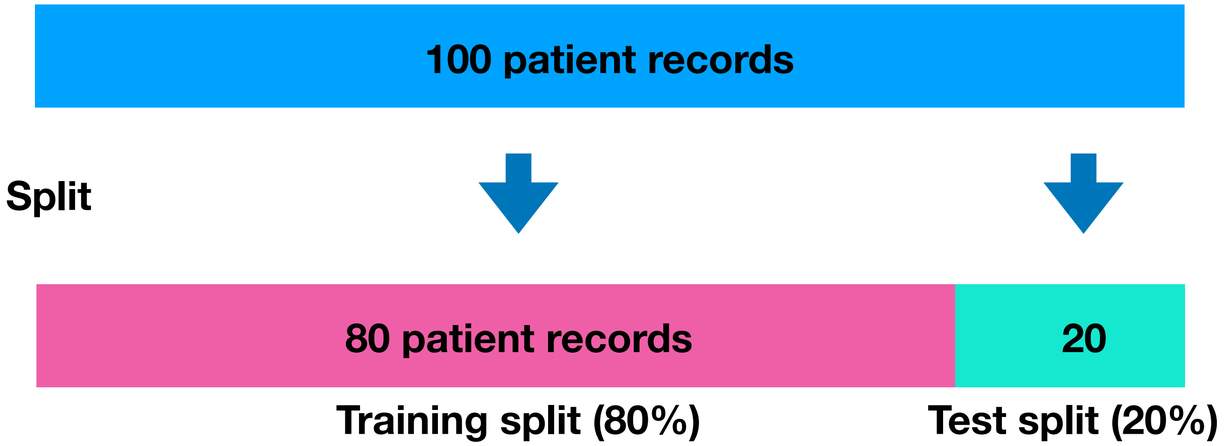






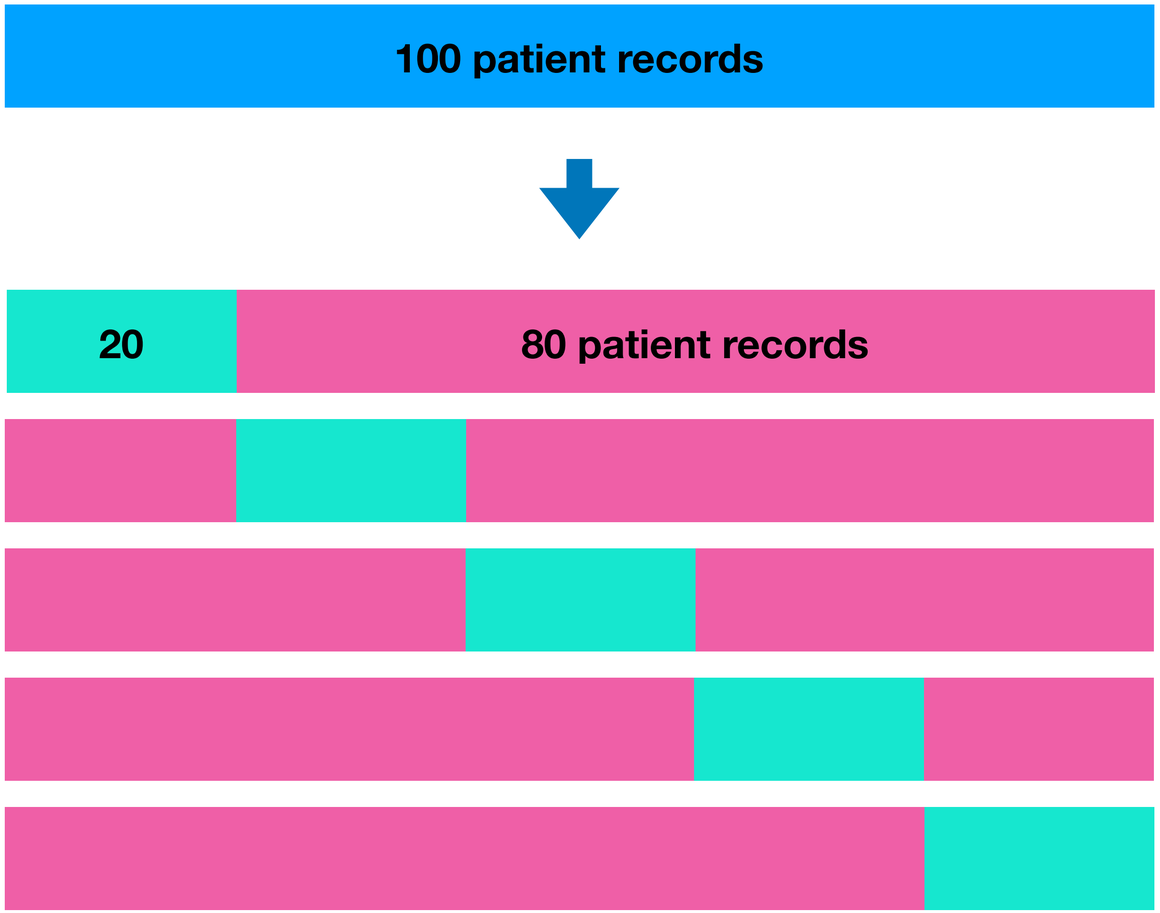
Cross-validation

Normal Train & Test Split



Model is trained on training data, and evaluated on the test data.

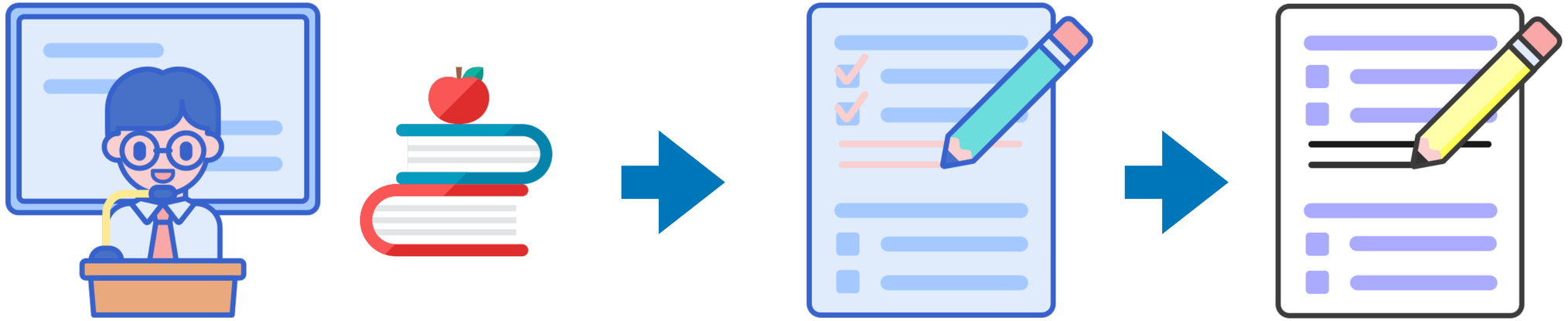
5-fold Cross-validation



Model is trained on 5 different versions of training data, and evaluated on 5 different versions of the test data.

The most important concept in machine learning

(the 3 sets)



Course materials (training set)

Practice exam (validation set)

Final exam (test set)

Generalization

The ability for a machine learning model to perform well on data it hasn't seen before.

Classification and Regression metrics

Classification	Regression
Accuracy	**R^2 (r-squared)**
Precision	Mean absolute error (MAE)
Recall	Mean squared error (MSE)
F1	Root mean squared error (RMSE)

Bold = default evaluation in Scikit-Learn

Which regression metric should you use?

- **R^2** is similar to accuracy. It gives you a quick indication of how well your model might be doing. Generally, the closer your **R^2** value is to 1.0, the better the model. But it doesn't really tell exactly how wrong your model is in terms of how far off each prediction is.
- **MAE** gives a better indication of how far off each of your model's predictions are on average.
- As for **MAE** or **MSE**, because of the way MSE is calculated, squaring the differences between predicted values and actual values, it amplifies larger differences. Let's say we're predicting the value of houses (which we are).
  - Pay more attention to MAE: When being $10,000 off is ***twice*** as bad as being $5,000 off.
  - Pay more attention to MSE: When being $10,000 off is ***more than twice*** as bad as being $5,000 off.