**2: Python for Machine Learning (6:10)**

### Learning Objectives

#### In this lesson you will learn about:

* + - Machine Learning applications
    - Python libraries for Machine Learning
    - Supervised vs Unsupervised Learning

1. Python is a popular and powerful **general-purpose programming language** that recently emerged as the preferred language among data scientists.
2. You can write your machine learning algorithm using python, and it works very well.
3. However, **there are a lot of modules and libraries already implemented in python that can make your life much easier.**
4. We try to introduce the Python packages in this course and use it in the labs to give you better hands-on experience.

**Python Libraries for Machine Learning (Numpy, SciPy, MatplotLib, Pandas, Scikitlearn)**



**NUMPY**

1. **The first package is Numpy, which is a math library to work with n-dimensional arrays** in Python.
2. It enables you to do computation efficiently and effectively.
3. It is better than regular python because of it’s amazing capabilities.
4. For example, for working with Arrays, Dictionaries, Functions, Datatypes, and working with Images, you need to know **Numpy**!

**SciPy:**

1. **SciPy is a collection of Numerical algorithms and Domain-Specific toolboxes, including signal processing, optimization, statistics and much more.**
2. SciPy is a good library for scientific and High-Performance Computation.

**Matplotlib:**

1. **Matplotlib is a very popular plotting package that provides 2D plotting as well as 3D plotting.**

**Basic Knowledge about these 3 packages, which are built on top of python, is a good asset for data scientists who want to work with real world problems.**

**PANDAS:**

1. **Pandas library, is a very High-Level Python Library that provides high-performance, easy to use Data Structures.**
2. **It has many functions for Data Importing, Manipulation and Analysis.**
3. **In particular, it offers Data Structures and operations for Manipulating Numerical Tables and Time Series.**

**SCIKIT-learn (aka SkLearn):**

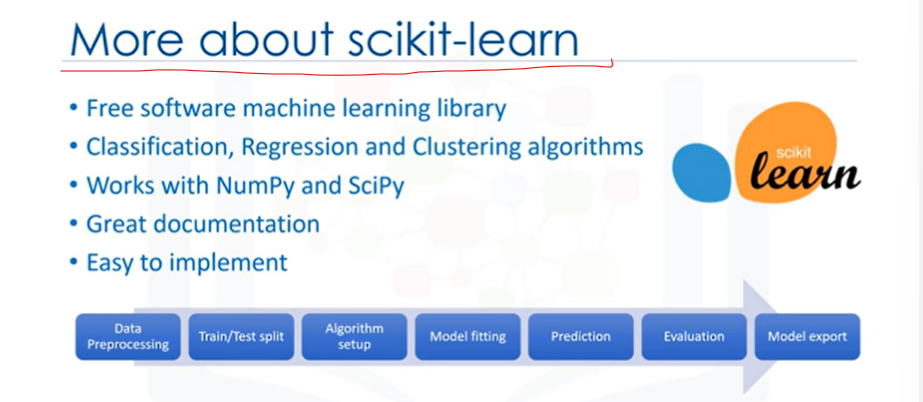
1. **Scikit-learn is a collection of Algorithms and Tools for ML, which is our focus here, and which you’ll learn to use within this course.**
2. As we’ll be using scikit-learn quite a bit, in the labs

**Scikit-learn**: **Explain more about it and show you why it is so popular among data scientists.?**

1. **Scikit-learn is a free machine learning library for the Python programming language.**
2. **It has most of the classification, regression and clustering algorithms, and it’s designed to work with the Python Numerical and Scientific libraries, NumPy and SciPy.**
3. Also, it includes very good documentation.
4. On top of that, implementing ML models with scikit learn is really easy, with a few lines of python code.
5. **Most of the tasks that need to be done in a ML pipeline are implemented already in scikit learn,**
6. **Including, Pre-processing of data, Feature selection, Feature extraction,**

**Train/Test Splitting, Defining the Algorithms, Fitting Models, Tuning Parameters,**

**Prediction, Evaluation, and Exporting the model.**



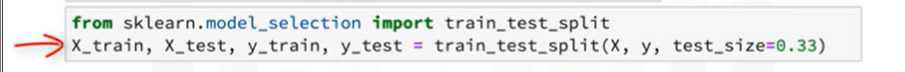
**Let me show you an example of how scikit learn looks like when you use this library.**

1. See how easily you can build a model with just a few lines of code.
2. Basically, ML Algorithms benefit from standardization of the data set.
3. If there are some outliers, or different scales fields in your data set, you have to fix them.
4. The preprocessing package of scikit learn provides several common utility functions and transformer classes to change raw feature vectors into a suitable form of vector for modeling.



1. **You have to split your dataset into train and test sets to train your model, and then test the model’s accuracy separately.**
2. Scikit learn can split arrays or matrices into random train and test subsets for you, in one line of code.

Then, you can setup your algorithm.



1. For example, you can build a classifier using a support vector classification algorithm.
2. We call our estimator instance clf, and initialize its parameters



1. Now, you can train your model with the train set.
2. By passing our training set to the fit method, the clf model learns to classify unknown cases.



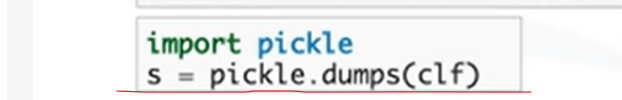
1. Then, we can use our test set to run predictions.
2. And, the result tells us what the class of each unknown value is.



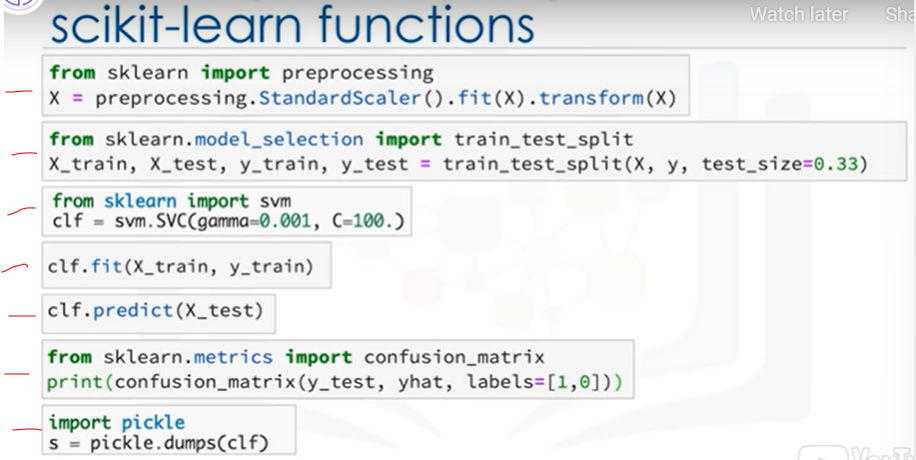
1. Also, you can use different metrics to evaluate your model accuracy, for example, using a confusion matrix to show the results.



1. And finally, you save your model.



**RECAP:**

****

**The most important point to remember is that the entire process of a Machine Learning task can be done simply in a few lines of code, using scikit learn.**

**Please notice that, though it is possible, it would not be that easy if you want to do all of this using Numpy or Scipy packages.**

**And of course, it needs much more coding if you use pure python programming to implement all of these tasks.**

**RECAP: SUMMARY**

**What have we learned? We have learned —**

**Python for Machine Learning (About)**

**Python Libraries for Machine Learning :**

* + **(Numpy, SciPy, MatplotLib, Pandas, Scikitlearn)**

**Scikit-learn**:

* + **Explain more about it and show you why it is so popular among data scientists.?**

**Let me show you an example of: CODING of ML Model Development & Saving in PKL**

* + **How scikit learn looks like when you use this library.(Coding)**