GATE WAY TO ML

SCIPY AND SCIKIT-LEARN

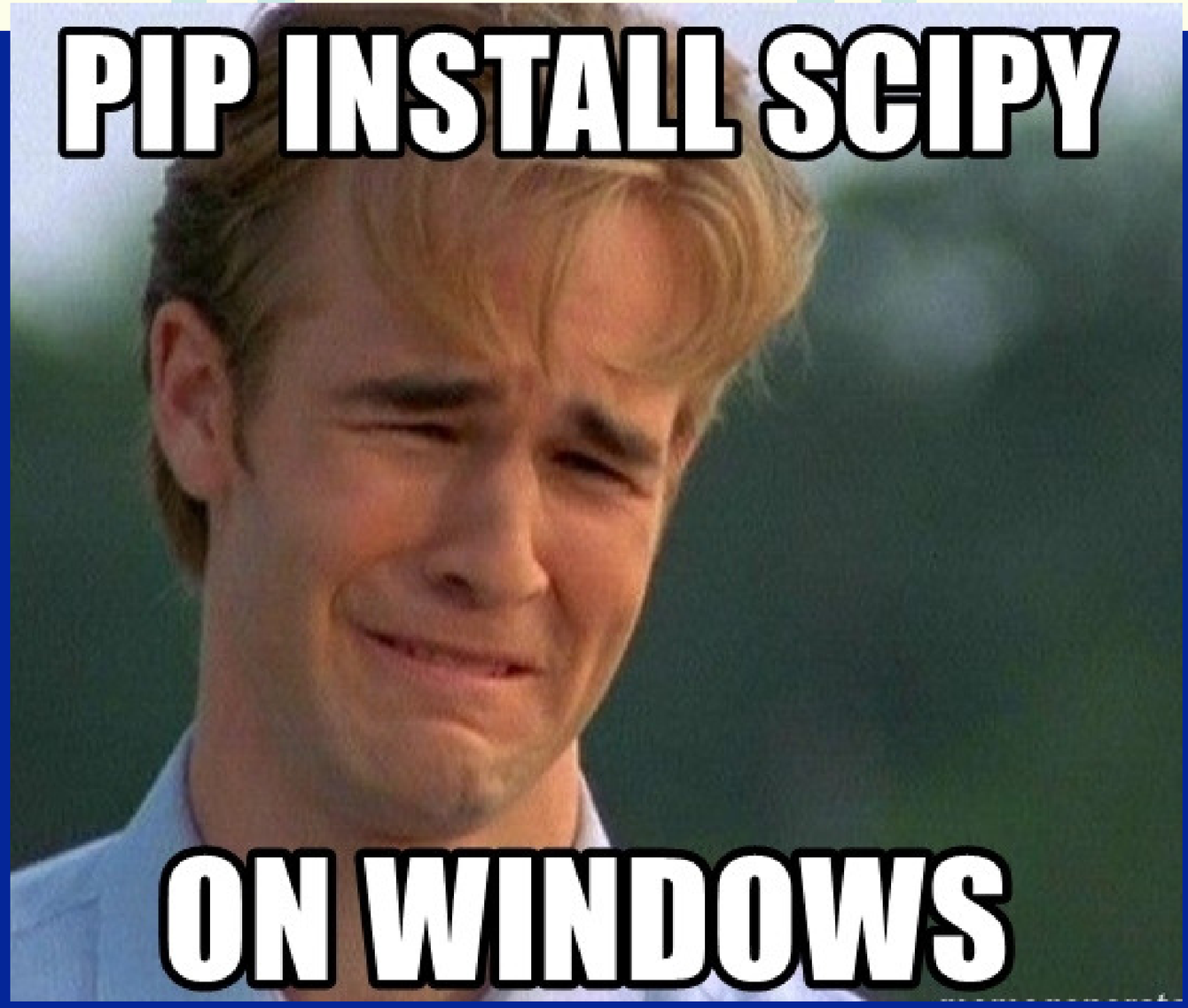
SCIPY IS AN ECOSYSTEM OF PYTHON LIBRARIES FOR MATHEMATICS, SCIENCE AND ENGINEERING. IT IS AN ADD-ON TO PYTHON THAT YOU WILL NEED FOR MACHINE LEARNING.

THE SCIPY ECOSYSTEM IS COMPRISED OF THE FOLLOWING CORE MODULES RELEVANT TO MACHINE LEARNING:

NUMPY: A FOUNDATION FOR SCIPY THAT ALLOWS YOU TO EFFICIENTLY WORK WITH DATA IN ARRAYS.

MATPLOTLIB: ALLOWS YOU TO CREATE 2D CHARTS AND PLOTS FROM DATA.

PANDAS: TOOLS AND DATA STRUCTURES TO ORGANIZE AND ANALYZE YOUR DATA.



INSTALLATION

CREATE ARRAY

NOTICE HOW WE EASILY CONVERTED A PYTHON LIST TO A NUMPY ARRAY.

```
 # define an array
import numpymy
list = [1, 2, 3]
myarray =
numpy.array(mylist)
print(myarray)
print(myarray.shape)
```

ACCESS DATA

ARRAY NOTATION AND RANGES CAN BE USED TO EFFICIENTLY ACCESS DATA IN A NUMPY ARRAY.

```
# access values
import numpymylist = [[1, 2, 3], [3, 4, 5]]
myarray = numpy.array(mylist)
print(myarray)
print(myarray.shape)
print("First row: %s") % myarray[0]
print("Last row: %s") % myarray[-1]
print("Specific row and col: %s") % myarray[0, 2]
print("Whole col: %s") % myarray[:, 2]
```

ARITHMETIC

NUMPY ARRAYS CAN BE USED DIRECTLY IN ARITHMETIC.

```
# arithmetic
import  numpy
myarray1 = numpy.array([2, 2, 2])
myarray2 = numpy.array([3, 3, 3])
print("Addition: %s") % (myarray1 +
myarray2)
print("Multiplication: %s") % (myarray1
* myarray2)
```

MATPLOTLIB

Matplotlib can be used for creating plots and charts. The library is generally used as follows:

- Call a plotting function with some data (e.g..plot()).
- Call many functions to setup the properties of the plot (e.g. labels and colors).
- Make the plot visible (e.g..show()).

LINE PLOT

THE EXAMPLE BELOW CREATES A SIMPLE LINE PLOT FROM ONE DIMENSIONAL DATA.

```
# basic line plot
import matplotlib.pyplot as pltimport
numpymyarray = numpy.array([1, 2, 3])
plt.plot(myarray)
plt.xlabel('some x axis')
plt.ylabel('some y axis')
plt.show()
```

SCATTER PLOT

BELOW IS A SIMPLE EXAMPLE OF CREATING A SCATTER PLOT FROM TWO DIMENSIONAL DATA.

```
# basic scatter plot
import matplotlib.pyplot as pltimport
numpyx = numpy.array([1, 2, 3])
y = numpy.array([2, 4, 6])
plt.scatter(x,y)
plt.xlabel('some x axis')
plt.ylabel('some y axis')
plt.show()
```

PANDAS

Pandas provides data structures and functionality to quickly manipulate and analyze data. The key to understanding Pandas for machine learning is understanding the Series and DataFrame data structures.

SERIES

A SERIES IS A ONE DIMENSIONAL ARRAY OF DATA WHERE THE ROWS ARE LABELED USING A TIME AXIS.

```
# series
import numpy
import pandas
myarray = numpy.array([1, 2, 3])
rownames = ['a','b','c']
myseries = pandas.Series(myarray, index=rownames)
print(myseries)
```

DATAFRAME

A DATA FRAME IS A MULTI-DIMENSIONAL ARRAY WHERE THE ROWS AND THE COLUMNS CAN BE LABELED.

```
# dataframe
import numpy
import pandas
myarray = numpy.array([[1, 2, 3], [4, 5, 6]])
rownames = ['a','b']
colnames = ['one','two','three']
mydataframe = pandas.DataFrame(myarray,
index=rownames, columns=colnames)
print(mydataframe)
```

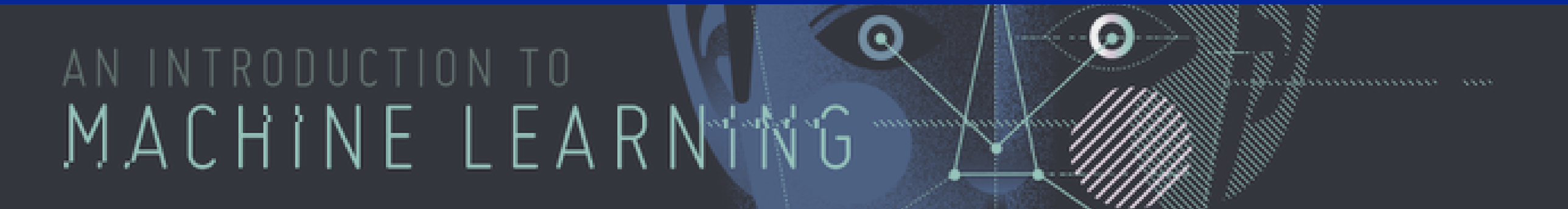
DATAFRAME

A DATA FRAME IS A MULTI-DIMENSIONAL ARRAY WHERE THE ROWS AND THE COLUMNS CAN BE LABELED.

```
print("method 1:")
print("one column:\n%s") % mydataframe['one']
print("method 2:")
print("one column:\n%s") % mydataframe.one
```

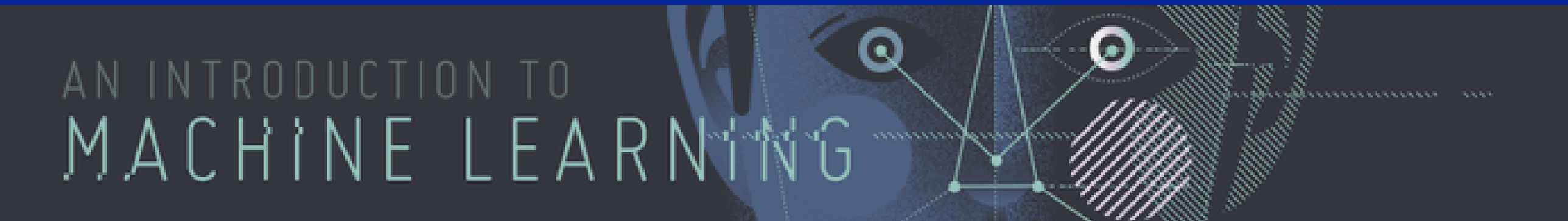
SCIKIT-LEARN

The scikit-learn library is how you can develop and practice machine learning in Python. It is built upon and requires the SciPy ecosystem. The name scikit suggests that it is a SciPy plug-inor toolkit. The focus of the library is machine learning algorithms for classification, regression,clustering and more. It also provides tools for related tasks such as evaluating models, tuning parameters and pre-processing data.



SCIKIT-LEARN

Like Python and SciPy, scikit-learn is open source and is usable commercially under the BSD license. This means that you can learn about machine learning, develop models and put them into operations all with the same ecosystem and code. A powerful reason to use scikit-learn.



Next Section

ML WORKFLOW