“Advanced Python Programming for Everybody”

Instructor: Ernest Bonat, Ph.D.
Senior Software Engineer
Senior Data Scientist
ebonat@15itresources.com
Cell: 503.730.4556

Module 7 Source Code

https://github.com/ebonat/intel_module_7

Module 7. "Python Data Ecosystem for Data Science Projects – Part 3"

What do you really need to know to become a Data Scientist?

- **Probability and Statistics** (undergraduate level)
- **Python Programming Language** (good level!)
- **Python Data Ecosystem** (good level!):

1. **NumPy** – fundamental package for scientific computing (Numerical Python - http://www.numpy.org/)
2. **pandas** – provides easy-to-use and high-performance data structures (https://pandas.pydata.org/)
3. **SciPy** - Python-based ecosystem of open-source software for mathematics, science, and engineering (https://www.scipy.org/)

4. **scikit-learn Machine Learning** – a simple and efficient tool for data mining and data analysis (http://scikit-learn.org/)
5. **matplotlib** – a 2D plotting library which produces publication quality figures in a variety of hard copy formats and interactive environments across platforms (https://matplotlib.org/)
6. **seaborn** - statistical data visualization (https://seaborn.pydata.org/)
7. **scikit-image** – a collection of algorithms for image processing (http://scikit-image.org/)

Practical Class

1. Tuning Hyperparameters Machine Learning Model
2. Machine Learning K-fold Cross Validation Model

"Train/Test Split and Cross Validation Python"

https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6