**Data Science with Python**

Sai Prasad Ashila

Dept. of Computer Science & Information Systems

Bradley university, Peoria, IL, USA

+1 (309) 868-3534

sashila@mail.bradley.edu

**ABSTRACT**

Python is an increasingly popular tool for data analysis. In recent years, a number of libraries have reached maturity, allowing R and Stata users to take advantage of the beauty, flexibility, and performance of Python without sacrificing the functionality these older programs have accumulated over the years. Python is a very powerful programming language used for many different applications. Over time, the huge community around this open source language has created quite a few tools to efficiently work with Python. In recent years, a number of tools have been built specifically for data science. As a result, analyzing data with Python has never been easier. Python is a general-purpose programming language that is becoming more and more popular for doing data science. Companies worldwide are using Python to harvest insights from their data and get a competitive edge. Unlike any other Python tutorial, this course focuses on Python specifically for data science.

**Categories and Subject Descriptors**

[**Programming Languages**]: Language Constructs and Features python, pandas, matplotlib, Numpy.

**1. INTRODUCTION**

Python has gathered a lot of interest recently as a choice of language for data analysis. I had [compared it against SAS & R](https://www.analyticsvidhya.com/blog/2014/03/sas-vs-vs-python-tool-learn/) some time back. Here are some reasons which go in favor of learning Python:

* Open Source – free to install
* Awesome online community
* Very easy to learn
* Can become a common language for data science and production of web based analytics products.

Needless to say, it still has few drawbacks too:

* It is an interpreted language rather than compiled language – hence might take up more CPU time. However, given the savings in programmer time (due to ease of learning), it might still be a good choice.

The differences between collaborative and content-based filtering can be demonstrated by comparing two popular music recommender systems – Last.fm and Pandora Radio.

**1.1 Python 2.7 v/s 3.4**

This is one of the most debated topics in Python. We will invariably cross paths with it, especially if we are a beginner. There is no right/wrong choice here. It totally depends on the situation and your need to use

**1.2 How to install Python?**

There are 2 approaches to install Python:

We can download Python directly from its [project site](https://www.python.org/download/releases/2.7/) and install individual components and libraries we need

Alternately, we can download and install a package, which comes with pre-installed libraries.

**2.** **Building a Predictive Model in Python**