

# Machine Learning Software Overview

## [Introduction - try to find some catchy title here...]

One of our key goals in this Machine Learning course is to give you a good deal of exposure to, and ultimately confident working knowledge of, some of the latest and best open-source software that either:

- directly implements complex Machine Learning algorithms; or
- indirectly facilitates obtaining, using and developing such Machine Learning software.

Given such a goal, we have decided to offer course content in not only one, but **two programming languages**, namely **R** and **Python**, the two leading, default go-to open-source software ecosystems for intensive, large-scale scientific computation in general and Machine Learning algorithms in particular.

*You have the option of doing your programming work in either language, and may get some “diligence bonus” when you use both!*

Below, we briefly discuss the key components of the R and Python ecosystems. You can notice a lot of parallelism between the two.

**Note:** the [Booth Analytics Club](#) has compiled a publicly-available [Learning Resources Catalog](#), including many online courses and software packages in R and Python. You may check it out every now and then.

## 1. The R Machine Learning Ecosystem

In R, we will work with the following:

- Programming Language: [R](#)
- Integrated Development Environment (IDE): [RStudio](#)
- Dynamic Documents: [R Markdown](#), already embedded in RStudio
- Default Package Repository: [Comprehensive R Archive Network \(CRAN\)](#)
- Pre-eminent Machine Learning Package: [Caret](#), a highly optimized wrapper around about 200 Machine Learning algorithms

Extras include:

- [Revolution Analytics](#)’ [doParallel](#) package for multi-core parallel computation

## 2. The Python Machine Learning Ecosystem

In Python, we will cover the following:

- Programming Language: [Python](#), particularly through the [Anaconda](#) pre-packaged distribution by Continuum Analytics
- Integrated Development Environment (IDE): [PyCharm](#)
- Dynamic Documents: [IPython Notebook](#)
- Default Package Repository: [Python Package Index \(PyPI\)](#)
- Pre-eminent Machine Learning Package: [SciKit-Learn](#), which gives you hundreds of Machine Learning algorithms as well as highly efficient data processing tools

Extras include:

- [DeepLearning.net](#)’s [Theano](#) package for fast numerical computation, especially beneficial for those of you whose machine comes with an NVIDIA graphics card.

### 3. Additional Supporting Software

Additionally, throughout the course you will also come across certain supporting tools that are useful for general open-source software development and distribution:

- **GitHub**: a web-based, revision-controlled, code hosting repository; our course materials will be distributed solely through our **GitHub course repository**
- **CygWin** (*for Windows users only*): a terminal to run Unix-style commands on Windows