

[HOME](#) [WORKSHOPS](#) [ESSAYS](#) [TALKS](#) [TEACHING](#) [QUOTES](#) [ABOUT](#)

Amit Kapoor

Crafting Visual Stories with Data



“All models are wrong, But some are useful.” – George Box

Applied Machine Learning

Every time we interact with an e-commerce site and see a recommendation to buy a product or we interact with our messenger app and see a chat bot in action, we are seeing machine learning in action. Strong mathematical theories underpin these machine learning application. And the Machine Learning library eco-system has matured to an extent that it is straight forward to write a few lines of code and have the ML back-end ready for one's application.

However, the challenge for many beginners is how to structure a business problem as a ML problem, and then go on to build, select and evaluate the right model. This workshop is designed to help learn how to apply machine learning to business problems. Real-life case studies are used to teach the various algorithms and techniques. The focus will be on applications, rather than on exposition of the various algorithms.

The workshop is divided into four major modules: Linear models, Model evaluation, Tree-based models and Model Selection. This is predominantly a hands-on course and will be 70% programming/coding and 30% theory.

Module 0: Introduction

- What is Machine Learning
- Types of ML: Supervised, Unsupervised, Reinforcement
- Types of ML problems: Regression, Classification

Module 1: Linear Models

- Linear Regression
- Logistic Regression

Module 2: Model Evaluation

- Training and Validation
- Model Evaluation Metrics - Accuracy, RMSE, ROC, AUC, Confusion Matrix, Precision, Recall, F1 Score
- Overfitting and Bias-Variance trade-off
- Regularization (L1/L2)
- K-fold Cross Validation

Module 3: Tree-based Models

- Decision Trees
- Bagging and Boosting
- Random Forest
- Gradient Boosting Machines
- Feature Importance

Module 4: Model Selection

- Model Pipelines
- Feature Engineering
- Ensemble Models (Advanced)
- Unbalanced Classes (Advanced)

Target Audience

- Anyone familiar with doing data analysis (using a scripting language like Python, R, SAS or programming languages like Java, Scala, C++) and wants to pick up the skills for machine learning.
- A programmer looking to transition in to building data driven products or a data scientist role.
- A beginner in data science with some experience in doing machine learning, but wants to get a deeper and a more applied perspective on using Machine Learning.

Pre-requisites

- Programming knowledge is mandatory. Attendee should be able to write conditional statements, use loops, be comfortable writing functions and be able to understand code snippets and come up with programming logic.
- Participants should have a basic familiarity of Python. Specifically, we expect participants to know the first three sections from this: <http://anandology.com/python-practice-book/>
- Participants should have experience with using Pandas and Jupyter Notebook. At the bare minimum, you should be able to understand and run the code in this [The Art of Data Science](#) repo. Refer to the Onion Notebook's and especially the Acquire, Refine, Transform and Explore sections.

Software Requirements

We will be using Python data stack for the workshop. Please install Ananconda for Python 3.5 for the workshop. That has everything we need for the workshop. For attendees more curious, we will be using Jupyter Notebook as our IDE. We will be using primarily scikit-learn libraries for most of the machine learning algorithms.

Facilitators' Profile

Amit Kapoor teaches the craft of telling visual stories with data. He conducts workshops and trainings on Data Science in Python and R, as well as on Data Visualisation topics. His background is in strategy consulting having worked with AT Kearney in India, then with Booz & Company in Europe and more recently for startups in Bangalore. He did his B.Tech in Mechanical Engineering from IIT, Delhi

and PGDM (MBA) from IIM, Ahmedabad. You can find more about him at <http://amitkaps.com/> and tweet him at [@amitkaps](#).

Bargava Subramanian is a practicing Data Scientist. He has 14 years of experience delivering business analytics solutions to Investment Banks, Entertainment Studios and High-Tech companies. He has given talks and conducted workshops on Data Science, Machine Learning, Deep Learning and Optimization in Python and R. He has a Masters in Statistics from University of Maryland, College Park, USA. He is an ardent NBA fan. You can tweet to him at [@bargava](#).

Interested to talk to me? [Schedule a time here](#).

Want updates on open events and books? [add your email](#)

[sign up](#)

Handcrafted by [Amit Kapoor](#) | [CC-BY-NC](#)

