Data Science Boot Camp

Presented by: Paul F. Roysdon, Ph.D.

Summary

This 2-week boot camp is for aspiring data scientists. We will take each student from an introduction to Jupyter Notebooks and Python, to the state-of-the-art in Deep Learning. Each section is organized with 30% lecture + 70% hands-on training. Each student will develop a uniquely tailored Capstone Project and present it at the end of the course.

Pre-boot camp Preparation

- Download and install Anaconda v1.9.7.
  - https://www.anaconda.com/distribution/
- Run test Jupyter Notebook "Welcome to DSBC" and report the numeric result on the first day of class.
  - https://github.com/raptor-ai/ds-bootcamp/notebooks/welcome.ipnyb

Time & Location

- September 23rd – October 4th, 2019
- 8:30am – 5:00pm, 30 min lunch @ 12:00pm, 15 min breaks @ 10:30am & 2:00pm.
- National Security Collaboration Center, North Paseo Building, Suite 4.212 (4th floor), University of Texas San Antonio.

Format

- 80 hours: 30% lecture + 70% Labs, Exercises and Demos
- 20 hours: Capstone Project (homework)
- 43 modules, with a quiz at the end of each module.
- Cumulative final exam based on quizzes. Passing score is 70% for JQS (beginner-level).

Prerequisites

- Interest in data science, data engineering, and programming.
- No experience necessary in data science. I will teach you what you need to know.
- Familiarity with Python and Microsoft Excel.

Requirements

- 100% attendance required to pass the course. No exceptions.
- Be 30 min early each day for security processing and check-in.
- Personal laptop to run Jupyter Notebooks on Google Collab.

Schedule

Day 1: Introduction to the Data Science

- AM: Introduction to Data Science, and Predictive Analytics, Anaconda and Python Programming.
- PM: EDA (Exploratory Data Analysis), Visualization, and Feature Engineering.

Day 2: Classification and Regression Algorithms

- AM: Linear algebra review. Evaluating Performance, Visualizing Features and Parameters.
- PM: Introduction to Predictive Modeling. Building a Classifier.

Day 3: Intro Machine Learning & Communicating Results

- AM: Statistics review. Intro to time series modeling, intro to Bayes Theorem.
- PM: Hypothesis testing, Communicating results with a customer.

Day 4: Unsupervised Learning

- AM: Calculus review, gradient descent. Dimension reduction (PCA, SVD, LSI).
- PM: Clustering (k-means, k-modes, hidden Markov models, expectation-maximization).

Day 5: Supervised Learning – Classification

- AM: k-neighbors, decision tree, naïve Bayes.
- PM: gradient boosting, bootstrapping, bagging, AdaBoost.

Day 6: Supervised Learning – Regression & Prediction

- AM: random forest, least squares (LS), weighted least squares (WLS).
- PM: linear and integer linear programming (LP and ILP), extended Kalman filtering (EKF).
- **Start Capstone Project.**

Day 7: Outlier Detection and Neural Networks

- AM: L1/L2 regression.
- PM: artificial neural networks (ANN).

Day 8: Deep Learning

- AM: Support vector machines (SVM), convolutional neural networks (CNN).
- PM: recurrent neural networks (RNN), long short-term memory (LSTM).

Day 9: Deep Learning

- AM: gated recurrent units (GRU), generative adversarial networks (GAN).
- PM: transformers.

Day 10: ML Ethics & Capstone Presentations

- AM: ML Ethics. Presentations.
- PM: Exam