Extended Data Science Bootcamp

Data Science Training Program
Booz | Allen | Hamilton
X Data Society

Advanced Classification

8 hours of instruction (Python)

Build upon your classification skills to develop advanced classification models and improve your predictive accuracy. Learn how to implement powerful algorithms, such as logistic regression and ensemble learning methods to identify the likelihoods of events and trends. You'll learn how to develop advanced classification models with increased accuracy, avoid pitfalls of classification models, and fine-tune parameters.

Syllabus & Topics Covered

1. Logistic regression

- Introduction to logistic regression and its relation to neural networks
- Assessing classifier performance
  (ROC curve, AUC, cutoff value selection)
- Tuning & regularization of logistic regression

2. Ensemble methods - random forests and gradient boosting

- Random forests
- Boosting methods
- Grid search and optimization of ensemble methods

Prerequisites

- Attendees must be comfortable using Python to manipulate data and must know how to create basic visualizations with ggplot2.
- Attendees must have a foundation in classification models and model accuracy measures.

Introduction to Text Mining

24 hours of instruction (Python)

Do you spend your days reading through reports and summaries? Do you want to be able to highlight key phrases and extract meaning automatically? This introduction to text mining provides the foundational skills you need in order to process, clean, and format text data for analysis. You'll learn how to integrate text mining into your work and extract key summary metrics and words from a corpus of documents.

Syllabus & Topics Covered

1. Preparing data for text mining

- What are text mining and Natural Language Processing (NLP)?
- Applications and intuition of text mining
- The process of cleaning and preparing text for analysis
- Working with different data formats – PDF, CSV, TXT

2. Visualizing text and understanding text distributions

- N-grams (bi-grams, tri-grams and quad-grams)
- Word clouds
- Histograms
- Summary metrics of corpora

3. Topic modeling within text

- Building a term document matrix
- Implementing bag-of-words technique on text data
- Building out TF-IDF
- Summary of topic modeling and implementing LDA
- Evaluating results and optimize number of topics

4. Text mining analysis

- Description of cosine similarity in text
- Building interactive network graphs to visualize similar documents
- Applying hierarchical clustering to text data
- Visualizing clustering as dendrogram and evaluate results

Prerequisites

Students must be comfortable using Python to manipulate data and must know how to create basic visualizations.
Additionally, students must have a foundation in classification models and model accuracy measures.

Building Scalable Models in Spark

8 hours of instruction

Learn how to optimize your code and to speed up current data processing using Spark. In this course, you will work through best practices of how and when to use Spark. You will explore what they can do with Spark and how to use distributed computing within Spark. Finally, you'll deploy a Spark application on AWS.

Syllabus & Topics Covered

1. Introduction to Spark and PySpark

- Working with data in PySpark
  - *RDDs vs. DataFrames vs. Datasets*
  - *When to use RDDs over DataFrames & Datasets*
- Features of Dataset in PySpark
  - *Optimized queries with Datasets*
  - *API for Datasets*

2. Introduction and overview of AWS for data science

- Overview the Data Analytics ecosystem within AWS
  - *Analytics*
  - *Data movement*
  - *Data lake*
  - *Predictive analytics and machine learning*

3. Working with ec2 in AWS

- Introduction and application of working with ec2
- Working with GPU using P2 instances

Prerequisites

Students must have experience with Python or another object-oriented language and foundational understanding of Spark.

AI: Neural Networks and Deep Learning

24 hours of instruction (Python)

Become an expert data scientist by implementing advanced techniques such as deep learning and neural networks. These state-of-the-art methods will allow you to build powerful predictive systems and find latent patterns in large amounts of data. This course dives into a host of algorithms known as artificial neural networks. You will learn the foundations of this complex and exciting topic, acquire practical skills to implement neural networks and deep learning models using TensorFlow and Keras libraries in order to solve real-world problems.

Syllabus & Topics Covered

1. Introducing the concept of neural networks

- Environment setup: TensorFlow 2.0
- Why TensorFlow?
- Introduction to neural networks
- Implement a neural network in Python using sklearn

2. Feed forward networks

- Forward and back propagation
- Single layer perceptron
- Implementation of models in TF

3. Best practices of model building – deep learning

- Batch learning
- GPU and when it is needed
- Data normalization
- Accuracy, precision, recall, F1 against "loss"

4. Auto encoders and Image classification - CNN

- Autoencoders and feature reduction
- Overview of CNNs and image data
- CNN architecture
- Simple CNNs – build and implement
- Optimizing CNN

Syllabus & Topics Covered Cont.

5. Recurrent Neural Networks (RNNs)

- RNN theory
- RNNs for time series data
- RNN implementation in TensorFlow
- Discuss best practices for time-series RNN models

6. Long short-term memory (LSTMs)

- Long Short-Term Memory (LSTM) theory
- LSTMs for text data
- LSTM implementation in TensorFlow

7. Reinforcement learning

- Reinforcement learning (RL) theory
- Implement RL using TensorFlow
- Expand CNNs to accomplish RNN tasks

Prerequisites

Any student in this program must have a strong foundation in Python and the common libraries: SciKit-Learn, Pandas, Numpy, and Matplotlib. Additionally, students should have a foundational knowledge of statistics, unsupervised machine learning algorithms, and classification algorithms.

Schedule

Tuesdays and Thursdays from 5pm - 7pm

Pricing

Payment: $2,900

Submit Your FlexEd Application

Apply for FlexEd Funding

You can apply for FlexEd funding. More info can be found here.

FlexEd Step-by-Step Guide

Use these FlexEd Instructions for a step-by-step guide to your application. The instructions and more key resources can be located in Degreed.

FlexEd Questions?

Visit the FlexEd FAQs here.