Data Analytics in Healthcare Workshop I

About the Course

Healthcare is now being transformed by data-driven discovery and prediction. Healthcare researchers are increasingly required to have skills necessary to perform data analytics at large-scale and know the complex ecosystem of tools and platforms in healthcare Big Data analytics.

These two workshops will cover the foundational topics in data science useful for healthcare professionals and give them skills to sample and apply the basic data analytics techniques. The topics covered are:

- Data Acquisition
- Data Management at Scale
- Data Analysis at Scale

Course Syllabus

Day 1:

Introduction and Data Acquisition

- Introduction to data science, history and context, technology landscape
- What is BigData
- The 7 Vs to know and Notions of Relevance, Utility, and Applicability
- Challenges in working with Large Volume
- Challenges in working with High Velocity
- Challenges in working with High Varieties- (Structured (tabular) , Semi-structured (JSON, graph networks), and Unstructured (text)
- What is “metadata?”
- Different types of Analytics
- The discipline of Healthcare Analytics
- Working with biomedical data: An overview of biomedical data types
- How to acquire the data and its challenges (API, Bio Repositories, and Scraping Hands on)

Day 2

Introduction to Data Processing and Distributed Computing

- Introduction to data processing at scale
- Introduction to relational databases
- Introduction to NoSQL
- Introduction to Data Scaling (Data warehouse, Data Marts, ETL Operations Data lake) and their challenges
- Introduction to distributed computing (Architecture Single Node vs. Distributed Storage)
- Introduction to the Grammar of Data and Structured Query Language (SQL) (Hands on)

Day 3:

Introduction to Large Scale Data Processing

- Scale out vs. Scale up
- Massively Parallel Processing and MapReduce
- Overview of various large scale processing architectures
- Introduction to distributed query processing
- Introduction to Hadoop ecosystem (HDFS, HIVE, etc)
- Writing MapReduce programs (hands on)

Day 4:

Introduction to High-Speed Distributed Computing

- What is High-speed distributed computing
- Basics of Indexing (Solr/Lucene/ElasticSearch)
- Introduction to Spark ecosystem
- Introduction to Spark SQL
- Working with Spark (hands-on)

Day 5:

Introduction to Image and Graph Processing

- Challenges in Storing and Processing of images
- Introduction to Genomic Data Science
- Introduction to Graph and Graph Databases
- Working with Images and Graphs processing (hands-on)

Data Analytics in Healthcare Workshop II

Building on the knowledge gained in the previous workshop, this workshop will cover the foundational skills to sample and apply the basic data analytics techniques. The topics covered are:

- Introduction to Natural Language Processing
- Introduction to Machine Learning
- Serving Data
- Healthcare analytics

Syllabus

Day 1:

Introduction to Natural Language Processing

- What is NLP and its application in healthcare data analytics
- Introduction to common NLP tools and frameworks
- Introduction to clustering texts
- Working with NLTK (hands-on)

Day 2 and 3:

- Introduction to Machine Learning

- Introduction to Exploratory Data Analysis
- Introduction to Statistical Models
- Introduction to Regression, Logistic Regression (hands-on scikit-learn)
- Introduction to Learning a Models Classification, SVM, Decision Trees and Random Forests
- Implementing Learning Models in Python (Hands-on)
- Introduction to Machine learning at scale

Day 4:

Image Processing and Serving Data

- Introduction to data visualization
- Data visualization in Python, Tableau, D3.js
- Datasets as a service
- Introduction to data serving at scale
- Image processing with Café

Day 5

Healthcare analytics

- Big Data Analytics for Healthcare
- Introduction to Genomic Data analytics
- Introduction to data quality
- Using data cleaning tools for high quality healthcare data
- Enriching datasets with web crawling and social networks data
- Working with real-world public datasets in healthcare (hands-on)