



AUGUST 13 - 17, 2018 / UNIVERSITY OF COPENHAGEN

BIG DATA ANALYSIS

— TOOLS AND METHODS

Big Data is omnipresent from industries to government and is frequently considered a completely new approach to problem solving. While the potential is often exaggerated, Big Data does indeed introduce new opportunities but also challenges. The ability to analyse and combine large amounts of data from different sources has obvious applications. However, the lack of quality in the data combined with a high variance means that conventional analysis often fails, while Machine Learning algorithms are less affected, if trained and used correctly.

This course will bring you to the forefront of the field by introducing you to the newest tools and methods in large-scale data analysis based on cutting-edge research and extensive experience.

WHAT YOU WILL LEARN

After the course, you will:

- Be able to set up basic Big Data Analysis from beginning to end: from retrieving and cleaning the data, to establishing the information level, extracting patterns and finding outliers, to curating the necessary data
- Be acquainted with a number of advanced tools like: Data cleaning, statistical methods for very large datasets, data stream analysis and finding patterns and outliers in Big Data, collecting data from instruments and devices (e.g. internet of things (IoT)) and hardware systems design for efficient BDA

COURSE CONTENT

Throughout the course, we will focus on using a few structured datasets which illustrate a commercial context and which will be used to demonstrate the different steps in Big Data Analysis.

Core elements:

- Data cleaning: Detecting and correcting (or removing) corrupt or inaccurate records
- Statistical methods: Robust methods for very large datasets and data with very large variance and outliers
- Finding patterns and outliers in Big Data: Which methods can be used to identify sparse patterns in very large datasets, and how can we identify data that does not follow the general pattern of a dataset?
- Collecting data from instruments and devices: How to collect, store, and analyze data from a multitude of sources (e.g. apparatus, IoT, etc.)
- Systems for Big Data Analysis: Common systems for BDA; Hadoop, PyDisco, etc., and hardware systems design for efficient BDA

Tools/methods introduced:

- Selected machine learning algorithms for large-scale data: Random forests, support vector machines, and large-scale exact nearest neighbour search
- Data curation: How to select data for long time curation, systems, techniques and standards for data curation

We will primarily be working with Python; however, all techniques that are covered are easily implemented with all standard data-analysis languages.



“Great course. I will recommend it to colleagues and friends.
Hope to come for another course at some point in the future”

Florian Friis, Lead Analyst, DONG Energy/Market Trading

PARTICIPANTS

The course is strictly focused on Big Data Analysis, thus a background in statistics and/or conventional data analysis is a prerequisite. This course assumes you have studied to at least Bachelor degree level and/or have several years of data analysis experience.

COURSE DATES

5 days, August 13 – 17, 2018, 9:00 – 16:30 at the University of Copenhagen, South Campus.

COURSE DIRECTOR

Troels C. Petersen, Associate Professor, Particle Physics, Niels Bohr Institute, University of Copenhagen

OTHER COURSE TEACHERS

Brian Vinter, Professor, eScience, Niels Bohr Institute, University of Copenhagen

Joachim Mathiesen, Associate Professor, Biocomplexity, Niels Bohr Institute, University of Copenhagen

COURSE FEE

EUR 2,680/DKK 19,900 excl. Danish VAT. Fee includes teaching, course materials and all meals during the course.



FOR MORE INFORMATION AND REGISTRATION:
copenhagensummeruniversity.ku.dk

“Good course. Thanks again!”

*Mahesh Kumar Murugaiah, Project Engineer,
Nord Sugar*