



Copenhagen  
Summer  
University



16 – 20 AUGUST 2021, 9:00 – 16:30



EUR 2,755 / DKK 20,500 EXCL. DANISH VAT

*Copenhagen Summer University*

# APPLIED MACHINE LEARNING AND BIG DATA ANALYSIS

Machine Learning is entering essentially all data-based fields, and Big Data is omnipresent from industries to governmental organizations. It is a new approach to problem solving, and while the potential is often exaggerated, Machine Learning 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.

## *“I have recommended this course to others already”*

Rasmus Berner, Finanstilsynet, Fuldmægtig, Applied Machine Learning 2020

This course will bring you to the forefront of the field of applied machine learning 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 a basic Machine Learning 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 for data cleaning, statistical analysis of very large datasets, data stream analysis, finding patterns and outliers in Big Data, collecting data from instruments and devices (e.g. internet of things (IoT)) and for hardware systems design for efficient handling and analysis of Big Data.

### COURSE CONTENT

Throughout the course, we will use examples of structured datasets, which illustrate a commercial context and which will be used to demonstrate the different steps in Big Data Analysis. Participants will also have the chance to ask questions about specific data and challenges.

#### Core elements

- Data cleaning and statistical methods: Detecting and correcting (or removing) corrupt or inaccurate records, and robust statistical methods for data with very large variance and cross checks.
- Machine Learning algorithms: Introduction to a variety of methods, how they work behind the scenes, their strengths and weaknesses, and their applications.
- 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: Hadoop, PyDisco, etc., and hardware systems design for efficient analysis.
- Selected machine learning algorithms for large-scale data:

Random forests, (deep) neural networks, support vector machines, and large-scale exact nearest neighbour search.

- 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.

### PARTICIPANTS

The course is strictly focused on Machine Learning and 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 16-20, 2021, 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

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

**Kenneth Skovhede**, Assistant Professor, X-ray and Neutron Science, Niels Bohr Institute, University of Copenhagen

### COURSE FEE

EUR 2,755/DKK 20,500 excl. Danish VAT. Fee includes teaching, course materials and all meals during the course.

## *For more information and registration*

 COPENHAGENSUMMERUNIVERSITY.DK	 <a href="mailto:csu@adm.ku.dk">csu@adm.ku.dk</a>	 +45 35 33 34 23
---	--	---

