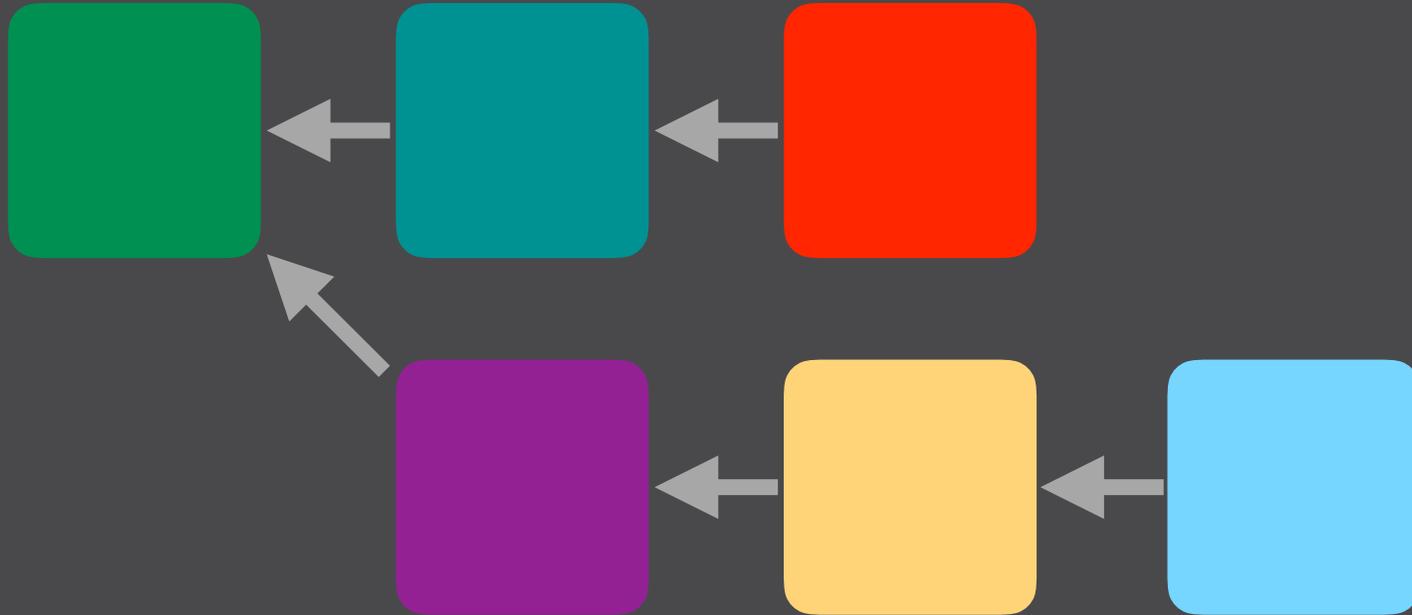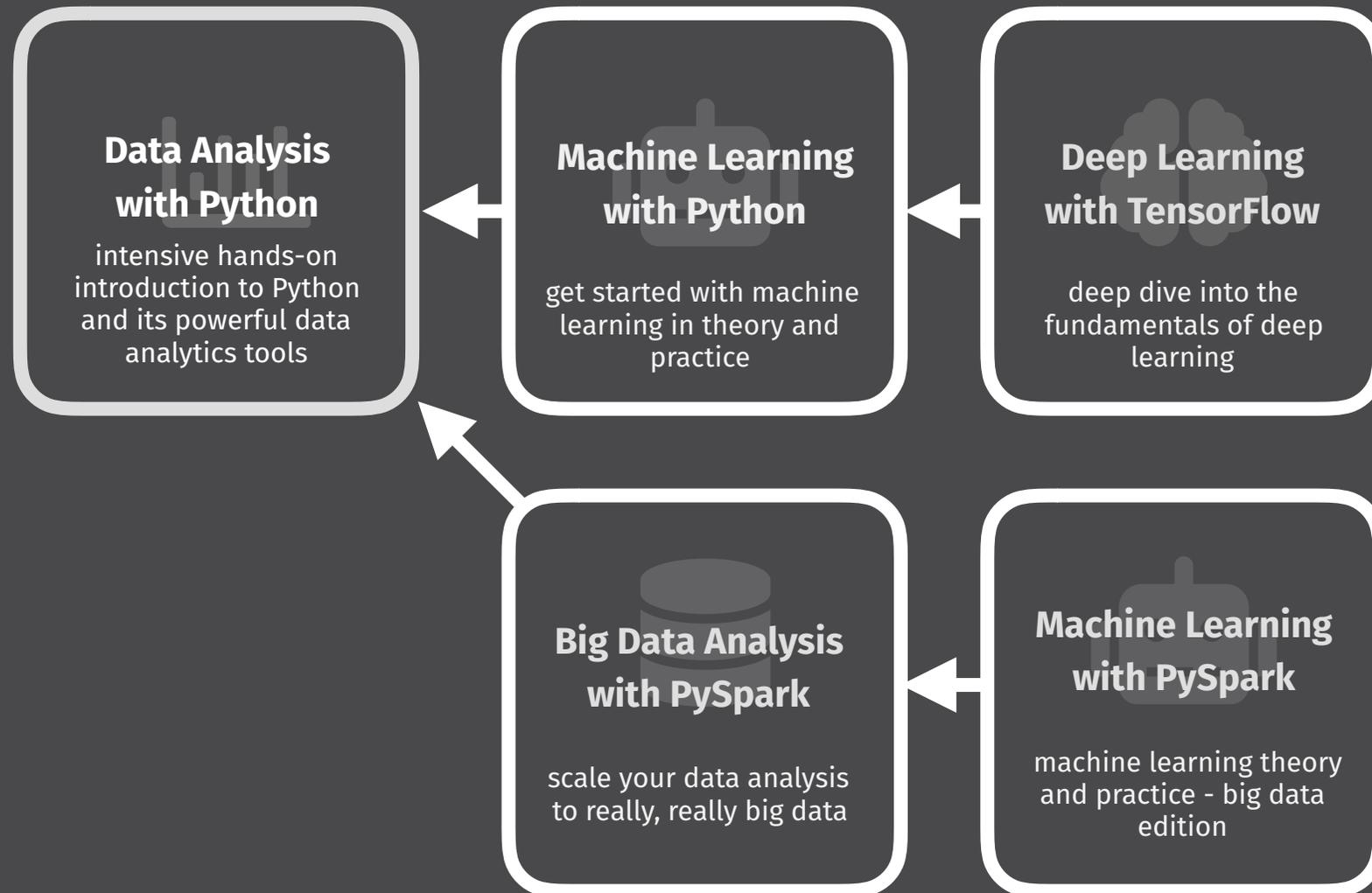# data science 101:
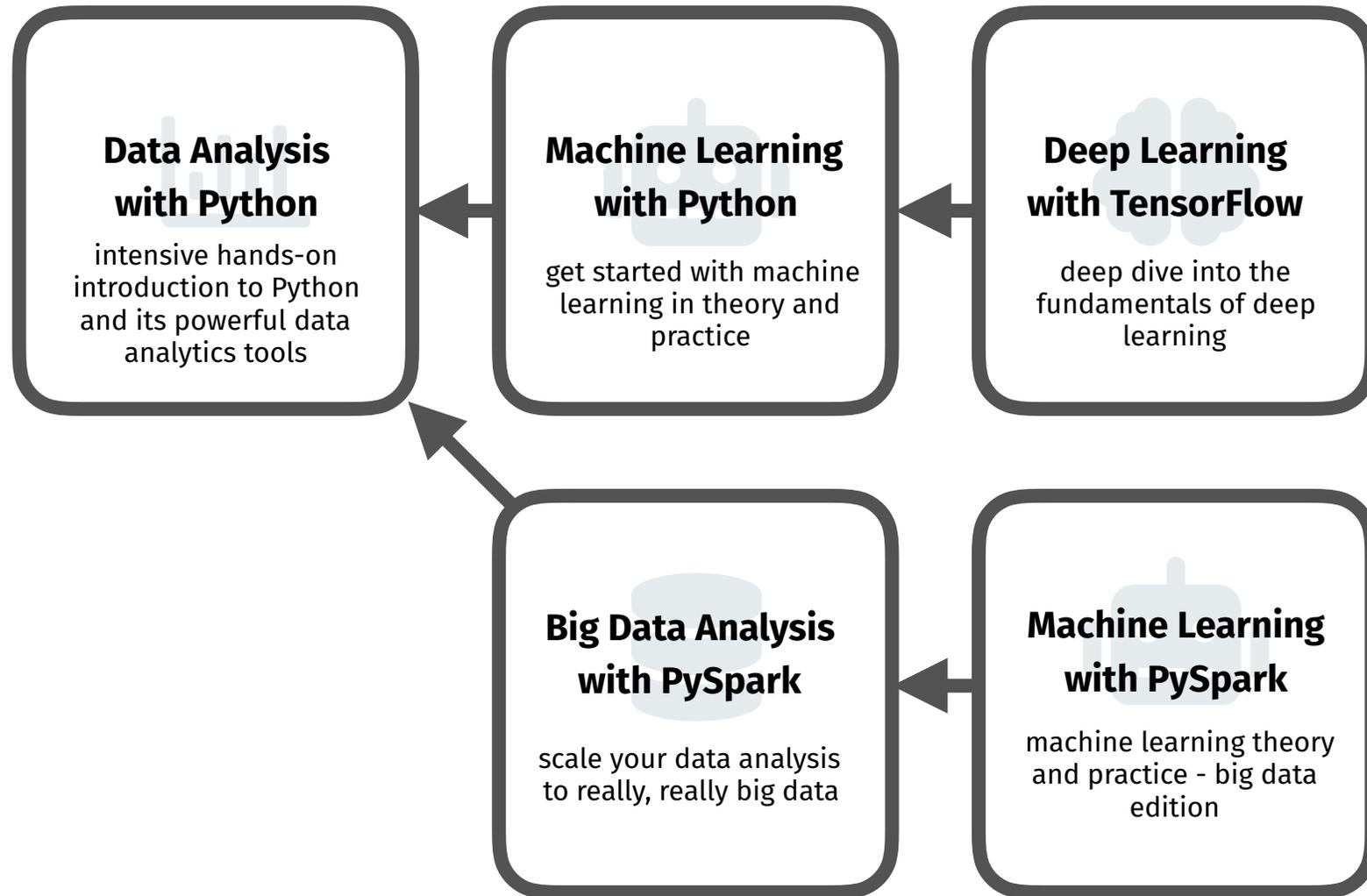## skill building blocks for the data-driven business

Dr. Christian Staudt · data scientist · in cooperation with **point8** data matters.

technical track
tools of the trade for the data scientist / analyst / application developer

Data Analysis with Python

intensive hands-on introduction to Python and its powerful data analytics tools

Machine Learning with Python

get started with machine learning in theory and practice

Deep Learning with TensorFlow

deep dive into the fundamentals of deep learning

Big Data Analysis with PySpark

scale your data analysis to really, really big data

Machine Learning with PySpark

machine learning theory and practice - big data edition

Data Analysis with Python

intensive hands-on introduction to Python and its powerful data analytics tools

Machine Learning with Python

get started with machine learning in theory and practice

Deep Learning with TensorFlow

deep dive into the fundamentals of deep learning

Big Data Analysis with PySpark

scale your data analysis to really, really big data

Machine Learning with PySpark

machine learning theory and practice - big data edition

← = "requires the knowledge from"

Setting

- up to 16 participants
- we provide a distraction-free workshop location

Didactics

- we aim for a healhy mix of lecture, demonstration, exercises (individual, pair programming and team exercises) and Q&A
- workshops are participatory - we integrate your questions and use cases

Material

- Jupyter Notebooks: interactive documents with instructions, code and graphics
- will be provided free of charge for self-study after the course

Language

- course language: English or German
- materials in English

**Data Analysis
with Python**

intensive hands-on
introduction to Python
and its powerful data
analytics tools

**Acronym: DAP
Level: Beginner
Duration: 2 days**

A 2-day intensive course covering the essentials for getting started with data analysis in Python. From the Python programming language to an ecosystem of powerful tools for data handling, analysis and visualization.

Curriculum (Part 1)

1. **Python Basics**
   Learn the basics of the Python programming language.

2. **Efficient Computing with numpy**
   Apply the numpy library to compute efficiently work large amounts of data.

3. **Data Handling with pandas**
   Learn to work with tabular data, supported by the pandas library.

4. **Plotting and Data Visualization**
   Visualize data with plots.

Data Analysis with Python

intensive hands-on introduction to Python and its powerful data analytics tools

Acronym: DAP
Level: Beginner
Duration: 2 days

A 2-day intensive course covering the essentials for getting started with data analysis in Python. From the Python programming language to an ecosystem of powerful tools for data handling, analysis and visualization.

Curriculum (Part 2)

1. **Introduction to Statistics**
    First steps with statistics concepts needed for data analysis.

2. **Handling Time Series with Pandas**
    Learn to work with time series data.

3. **Outlook: Machine Learning**
    A preview on machine learning applications.

Machine Learning with Python

machine learning theory and practice

Acronym: MLP
Level: Advanced
Duration: 2 days

A 2-day intensive course enabling the participants to build machine learning applications with Python and the scikit-learn framework. We work on typical machine learning cases using hands-on examples.

Curriculum (Part 1)

1. **Introduction to Machine Learning**
   An overview over the field of machine learning.

2. **ML for Classification**
   Build a classification model and learn about the building blocks of ML with Python

   A. **About Classification**
      Learn about classifiers and how to measure the quality of their decisions.

   B. **Feature Engineering**
      How to build better features.

   C. **Algorithm Selection and Hyperparameter Tuning**
      Select the right algorithm for the job and optimize performance.

   D. **Exercise: Classifier**
      Build your own classification pipeline from scratch.

Machine Learning with Python

machine learning theory and practice

**Acronym: MLP
Level: Advanced
Duration: 2 days**

A 2-day intensive course enabling the participants to build machine learning applications with Python and the scikit-learn framework. We work on typical machine learning cases using hands-on examples.

Curriculum (Part 2)

3. **ML for Regression**
    Learn about regressors and how to measure the quality of their prediction.

    A. **Exercise: Predicting House Prices**
        Apply regression to predict house prices.

4. **Building a Recommender Engine**
    Use machine learning to generate movie recommendations.

5. **Unsupervised Learning: Clustering**
    Apply clustering algorithms to detect structure in the data.

    A. **Exercise: Clustering**
        Practice on a clustering problem.

Big Data Analysis with PySpark

scale your data analysis to really, really big data

Acronym: BDAS
Level: Advanced
Duration: 1 day

This course gets you started with scaling data analysis methods to large amounts of data, using PySpark to control distributed computing on a cluster.

Curriculum

1. **Processing Big Data**
   What strategies do we have available to compute efficiently with increasing amounts of data? What is a cluster, and when do we need one?

2. **Spark Fundamentals**
   An overview of Spark - a framework for programming distributed computation, using PySpark, its Python API - core data structures and operations.

3. **Submitting Spark Jobs**
   How to submit jobs to a Spark cluster for batch processing.

4. **Spark and Structured Data**
   Working with structured data in Spark.

Machine Learning with PySpark

machine learning theory and practice - big data edition

**Acronym: MLS
Level: Advanced
Duration: 2 days**

A 2-day intensive course enabling the participants to build big data machine learning applications with PySpark. We work on typical machine learning cases using hands-on examples.

Curriculum (Part 1)

1. **Introduction to Machine Learning**
   An overview over the field of machine learning.

2. **ML for Classification**
   Build a classification model and learn about the building blocks of ML with PySpark

   A. **About Classification**
      Learn about classifiers and how to measure the quality of their decisions.

   B. **Feature Engineering**
      How to build better features.

   C. **Algorithm Selection and Hyperparameter Tuning**
      Select the right algorithm for the job and optimize performance.

   D. **Exercise: Classifier**
      Build your own classification pipeline from scratch.

Machine Learning with PySpark

machine learning theory and practice - big data edition

Acronym: MLS
Level: Advanced
Duration: 2 days

A 2-day intensive course enabling the participants to build big data machine learning applications with PySpark. We work on typical machine learning cases using hands-on examples.

Curriculum (Part 2)

3. **ML for Regression**
   Learn about regressors and how to measure the quality of their prediction.

   A. **Exercise: Predicting House Prices**
      Apply regression to predict house prices.

4. **Building a Recommender Engine**
   Use machine learning to generate movie recommendations.

5. **Unsupervised Learning: Clustering**
   Apply clustering algorithms to detect structure in the data.

   A. **Exercise: Clustering**
      Practice on a clustering problem.

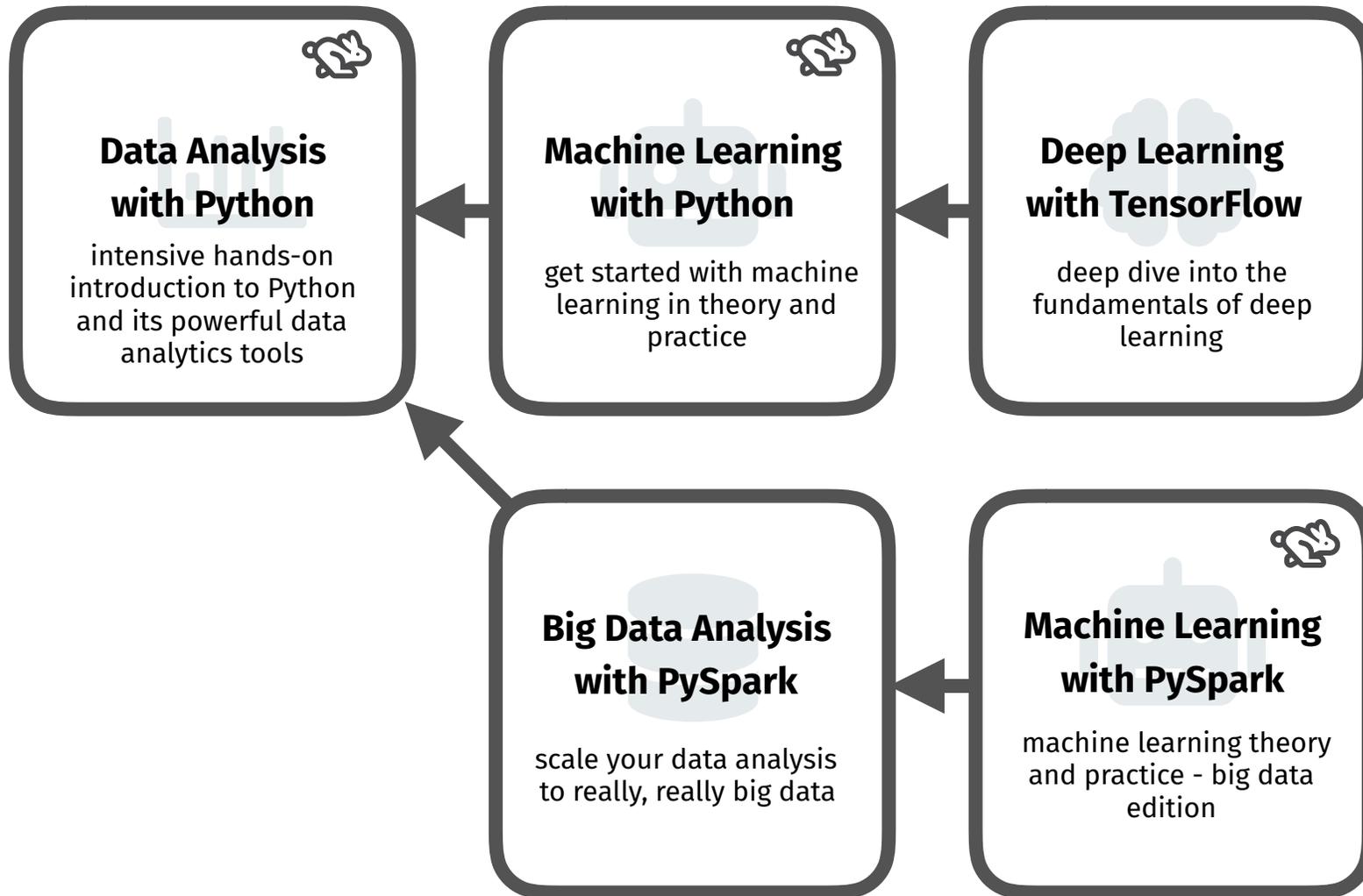Deep Learning
with TensorFlow

deep dive into the
fundamentals of deep
learning

Acronym: DLT
Level: Advanced
Duration: 1 days

This course is an introduction to deep learning with the TensorFlow framework for those looking to deepen their machine learning skills.

Curriculum

1. **Our First Neural Network for Digit Recognition**

2. **Keras: A High-Level API for Building Networks**

3. **TensorFlow Fundamentals**

4. **TensorBoard: Watch your Network**

5. **More Image Recognition: Cat or Dog?**

6. **Building an Autoencoder**

Data Analysis with Python

intensive hands-on introduction to Python and its powerful data analytics tools

Machine Learning with Python

get started with machine learning in theory and practice

Deep Learning with TensorFlow

deep dive into the fundamentals of deep learning

Big Data Analysis with PySpark

scale your data analysis to really, really big data

Machine Learning with PySpark

machine learning theory and practice - big data edition

We also offer an **express version** of the 2-day courses, condensing them to **one intense day of training**.

Building skill takes time. We recommend this only for participants…
- … with prior experience in programming, data analysis or machine learning with comparable tools (e.g. *R*)
- … attending several modules in a row (e.g. DAP -> MLP -> DLT)

Dr. Christian Staudt
Freelance Data Scientist

Technical Track Instructor Bio

- **Independent Data Scientist**
  July 2016 – now
  data science contract roles in multiple industries - analytics applications, big data solutions, machine learning, artificial intelligence

- **Doctoral Researcher at Karlsruhe Institute of Technology (KIT)**
  October 2012 – June 2016
  algorithms, parallelism, data analysis, software engineering

Education

- Doktor der Naturwissenschaften (Dr. rer. nat.) (= PhD), Computer Science, Karlsruhe Institute of Technology (KIT)

- Diplom (= Master's Degree), Computer Science, 2005 – 2012, Karlsruhe Institute of Technology (KIT)

management track
getting data-driven innovation done

point8
data matters.

point8
data matters.

Experimentelle Physiker mit Innovationsdrang, interessiert an Zukunftstechnologien und immer auf der Suche nach neuen Herausforderungen. Das Point 8 Team bringt Big Data Know-how vom CERN in die Wirtschaft und unterstützen Unternehmen und Organisationen mit Datenanalyse, Machine Learning und Simulationen.

Point 8+

Data Science
Case Workshop

discover the case for data
science in your enterprise

Data Science
Mission Control

mastering data-driven innovation - from business
case to production

Data Science Case Workshop

discover the case for data science in your enterprise

Acronym: DSCW

We conduct a Data Science Case Workshop with our clients to review the value of the enterprise's data and discover business cases for data science methods.

Agenda

1. **Introduction**

2. **Business Case Discovery**

3. **Data Assessment**

4. **Hands-on Data Exploration**

5. **Business Case Analysis**

6. **Technical Feasibility Analysis**

7. **Results Workshop**

Data Science Mission Control

Big Data, Machine Learning... Rocket Science?

Unser Konzept: Datengetriebene Innovation meistern - vom Business Case bis zur Produktivumgebung

realistisch: gegen den Hype, für das echte Potential von Big Data, Machine Learning und Artificial Intelligence

kollaborativ: die wichtigsten Stakeholder aus Ihrem Unternehmen an einem Tisch - von Management bis IT

integriert: technische Grundlagen im Blick ebenso wie zwischenmenschliche Kommunikation

fokussiert: Intensivworkshop in ablenkungsfreier Umgebung

fundiert: das Team der Data Scientists von Point 8 vermittelt Kompetenzen aus der langjährigen Praxis in Grundlagenforschung und Industrieprojekten

point8
data matters.

Hype Management

Lernen Sie zu trennen zwischen dem Hype und dem echten Potential von Big Data, Machine Learning und Artificial Intelligence.

Stakeholder Cooperation

Gestalten Sie die Zusammenarbeit zwischen den Stakeholdern eines Data Science-Projektes erfolgreich.

Best Practices & Processes

Etablieren Sie Prozesse und Workflows, die erfolgreiche Data Science-Projekte ausmachen.

Data Science Mission Control

Big data, machine learning... rocket science?

Mastering data-driven innovation - from business case to production

realistic: against the hype, for the real potential of big data, machine learning and artificial intelligence

collaborative: bringing the important stakeholders in your business to the table - from IT to management

integrated: considering technical as well as human factors

focused: an intensive workshop in a distraction-free environment

informed: the team of data scientists of Point 8 conveys competence from years of experience in research and industry

point8
data matters.

Hype Management

learn to distinguish hype and true potential around big data, machine learning and artificial intelligence

Stakeholder Cooperation

shape successful cooperation between the stakeholders of a data science project

Best Practices & Processes

establish processes and workflows, that are the foundation of successful data science projects

Ablauf

Kick-Off
Zu Beginn unseres Workshops starten wir mit einer Keynote Lecture und einem ersten Kick-Off zum Kennenlernen.

Hype Management
Lernen Sie zu trennen zwischen Hype und echtem Potential von Big Data, Machine Learning und Artificial Intelligence.

Stakeholder Cooperation
Gestalten Sie die Zusammenarbeit zwischen den Stakeholder eines Data Science-Projektes erfolgreich.
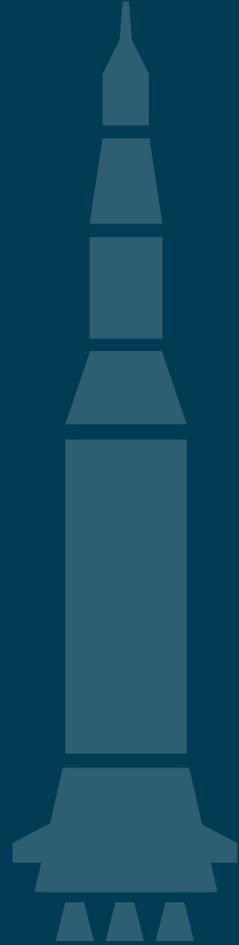
Best Practices & Processes
Etablieren Sie Prozesse und Workflows, die erfolgreiche Data Science-Projekte ausmachen.

Let's start and do it!
Das Licht am Horizont, Potentiale im Unternehmen erkennen und die ersten Schritte planen und angehen.

Agenda

Kick-Off
We start our workshop with a keynote lecture and an opportunity for getting to know each other

Hype Management
Learn how to separate the hype from the real potential of big data, machine learning and articial intelligence.

Stakeholder Cooperation
shape successful cooperation between the stakeholders of a data science project

Best Practices & Processes
establish processes and workflows, that are the foundation of successful data science projects

Let's start and do it!
Recognize potential in your enterprise and go for actionable steps.

our track record

We continually update and improve **data science 101** according to our experience with previous workshops and industry data science projects. Together, we have so far conducted **over 30 successful workshops for industry leaders in the machine building, telecommunications and energy sectors.**

References (public)

get in touch

…. for bookings and inquiries about data science 101, including exclusive and custom workshops. **Current courses are listed at: https://point-8.de/trainings.html**

Dr. Christian Staudt
Freelance Data Scientist
Hamburg

email: **mail@clstaudt.me**

web: **http://clstaudt.me**

twitter: **@C_L_Staudt**

Data Science as a Service
Dortmund

email: **kontakt@point-8.de**

web: **http://point-8.de**

twitter: **@point8hq**