

Category: Blockchain / AI / ML

AI Data Science #020101

Das Training findet sich täglich vom 08:30 bis 12:30 statt.

Trainings-Dauer: 14.04.2025 bis zum 13.06.2025 (40 Tage)

Feiertage: 18.04, 21.04, 01.05, 29.05, 09.06

Die ersten 3 Tage finden sich am Ort in Circle 6 am Zürich Flughafen statt.

Das Training in den Resttagen (bis zum 13.06.2025) findet sich nur Online statt.

The minimum hardware requirements for Training Students:

- 1 GHz single-core processor,
- 2 GB of RAM,
- 10 GB of free disk space,

Provided Compute Resources in Learning Environment

- SSH / SFTP / SCP access to Training-Partner remote web-servers for students
 - User accounts for each student set up
 - 1 Gb RAM and 10 Gb file storage provisioned for each student
- Github Org including all lecture materials
- Heroku / AWS / Azure organization including sufficient space for students to deploy applications
- 5-day GPU Access on Digi2Cloud Paperspace distribution (for final projects -- only in the case of Neural Network projects)
- Any / All additional 3rd party APIs for data access.
- Any / All 3rd party software for project development.
- Slack organization provisioned for each bootcamp
- Jupyter Hub

Prior Knowledge Requirements / Anforderungen

- At 1 year programming in Python.
- Solid understanding of basic mathematics (algebra, functions)
- Prior knowledge of Unix / Bash is helpful (not mandatory)
- Prior knowledge of SQL or relational databases is useful (not mandatory)



Key Learnings

- Programming for Data Science:
 - Pandas
 - Numpy
- SQL and Object Relational Databases
- Data Structures & Algorithms
- Graphing and Visualization:
 - Plotly
 - Dash
- Statistics for Data Science
 - Hypothesis Testing
 - Descriptive Statistics
 - Alpha, Beta, Confidence Intervals
- Probability Distributions and Samples
- Statistical Inference
- Forecasting
- Linear Regression
- Non-Linear Regression
- Time Series
 - Decomposition
 - Regression techniques
 - Forecast evaluation
- Data Manipulation:
 - Imputation
 - Fill methods
 - Data Cleaning and Aggregation
- Machine Learning with Python
 - Decision Trees
 - Clustering Algorithms
 - Model Evaluation and Cross Validation
 - Feature Engineering and Selection
 - Random Forest
 - Dimensionality Reduction (PCA, KPCA)
- Natural Language Processing
 - Large Language Models
 - Retrieval Augmented Generation
 - Vector Search and Nearest Neighbor Search
- Data Visualization
 - Basic Principles of Data Visualization
 - Intensive Data visualization techniques (Dash)

Inhalt

Wk. 1	
Mon	(08:30) 1 hour lecture on Python, Git, and version control. (09:30) 30 min Practice with Python and Git & 30 min solutions. (10:30) 1 hour (coding together) tutorial: Setting up SSH Keys, the Project Repository, and exercise repository. (11:30) Independent work on Homework, reading course materials and "warm-up" exercises.
Tue	(08:30) 1 hour Introduction Python, Git, Unix., Bash (09:30) 30 min setup of git repo on remote and local servers. (10:30) 1 hour (coding together) tutorial on Git Flow and Management of Merge Conflicts. (11:30) Independent work on Homework and exercises.
Wed	(08:30) Pandas & Numpy Lecture (09:30) 30 min exercises on Pandas (10:30) 1 hour (coding together) tutorial in advanced Pandas, introduction to Matplotlib (11:30) 1 hour Home exercises and coding practice.
Thur	(08:30) 1 hour lecture on major data types, and treatment of various data types. (09:30) 30 min exercises & 30 min solutions and explanations on the topic of Data types (10:30) 1 hour (coding together) tutorial on advanced pandas (merge, melt, concat) (11:30) 1 hour Home exercises and coding practice.
Fri	(08:30) 1 hour lecture on more advanced data visualization. (09:30) 30 min exercises & 30 min solutions and explanations on the topic of advanced data visualization (10:30) 1 hour (coding together) tutorial advanced data visualization and data analysis (11:30) 1 hour Home exercises and coding practice.



Wk. 2	
Mon	(08:30) 1 hour lecture on desc. statistics and hypothesis testing (09:30) 1 hour exercises and solutions on data-driven hypothesis generation and descriptive statistics (10:30) 1 hour (coding together) workshop: Generating descriptive statistics from fresh datasets (11:30) 1 hour Project Work and Questions
Tue	(08:30) Introduction to Monte-Carlo random sampling (09:30) 30 min exercises on monte carlo sampling (10:30) 1 hour (coding together) comparing random samples from different distributions (11:30) 1 hour Project Work & Questions
Wed	(08:30) 1 hour lecture Conditional Distributions (09:30) 30 min exercises & 30 min solutions and explanations & 1 hour (coding together) tutorial on authentication & User Management (10:30) 1 hour (coding together) tutorial on implementing conditional sampling (11:30) 1 hour Project Work & Questions
Thur	(08:30) (lecture) Linear Regression (continued) (09:30) 30 min exercises & 30 min solutions of exercises in multiparameter linear regression (10:30) 1 hour (coding together) tutorial on implementing linear regression estimation for large datasets (11:30) 1 hour Project Work & Questions (Add Auth to App!)
Fri	(08:30) (lecture): Non-Linear Regression (continued) (09:30) 1 hour exercises + solutions for non-linear regression techniques on large datasets (10:30) 1 hour (coding together) tutorial on building non-linear regression forecasting pipeline (11:30) 1 hour Project Work & Questions



Wk. 3	
Mon	(08:30) Normalization, Imputation and Data Cleaning (09:30) Exercises in data cleaning: how fast can you clean your dataset? (10:30) 1 hour (coding together) tutorial: Working with massive datasets (11:30) Independent work on Projects
Tue	(08:30) Advanced data transformations for addressing multicollinearity (09:30) Exercises in multicollinearity: when to transform? (10:30) 1 hour (coding together) advanced data imputation (11:30) 1 hour Project Work & Questions
Wed	(08:30) Dimensionality Reduction + Clustering (09:30) 1 hour exercises PCA, kPCA (10:30) 1 hour (coding together) tutorial on dimensionality reduction and clustering (11:30) 1 hour Project Work & Questions
Wed	(08:30) 1 hour lecture on Time-Series data: How is it different? (09:30) 1-hour exercises on manipulation of dependant datasets (10:30) 1 hour (coding together) tutorial on timeseries datasets in python (pandas) (11:30) 1 hour Project Work & Questions
Fri	(08:30) 1 hour lecture on Time-Series regression (09:30) 1-hour exercises ARM and ARIMA regression models (10:30) 1 hour (coding together) tutorial on timeseries datasets in python (pandas) (11:30) 1 hour Project Work & Questions

Wk. 4	
Mon	(08:30) Lecture on Assessing Classification Accuracy (09:30) 30 min exercises + 30 min solutions Classification Accuracy (Predicted vs. Observed, ROC, MSE) (10:30) 1 hour (coding together) tutorial: Setting up a classification pipeline (11:30) 1 hour project work and questions
Tue	(08:30) 1 hour lecture on Random Forest (09:30) 30 min exercises + 30 min solutions in using the Random Forest Algorithm (Machine Learning) (10:30) 1 hour (coding together) tutorial: Preparing data for Random Forest (11:30) Project work and questions
Wed	(08:30) 1 hour lecture on Bagging and Ensemble methods (09:30) 30 min exercises + 30 min solutions on Machine Learning: how to use machine learning methods ? (10:30) 1 hour (coding together) tutorial: Setting up a machine learning model (11:30) 1 hour project work and questions
Thur	(08:30) Data Structures and Algorithms (09:30) 1 hour data-structures & algorithms (10:30) 1 hour (coding together) tutorial on Data Structures and Algorithms (Technical Interview Prep) (11:30) 1 hour Project Work & Questions
Fri	(08:30) Technical Interview Drilling (all participants practice job interviews + reviewing interviews) (11:30) 1 hour Project Work & Questions

Wk. 5	
Mon	(08:30) 1 hour lecture on Web-Development and Dash (09:30) 30 min exercises + 30 min solutions in Dash operations (10:30) 1 hour (coding together) tutorial: Setting up a Dash Application and visualizing a graph (11:30) Setup first deployment (Heroku)
Tue	(08:30) 1 hour lecture on Plotly and advanced Dash Callbacks (09:30) 30 min exercises in Dash data visualization (10:30) 1 hour (coding together) tutorial: integrating statistical algorithms with a Dash Application (11:30) Setup a second deployment (Heroku)
Wed	(08:30) Introduction to REST and APIs (09:30) 30 min exercises & 30 min solutions and explanations with Integrating REST api into pipeline (10:30) 1 hour (coding together) tutorial - Adding API data to our Dash app (11:30) 1 hour Project Work & Questions
Thur	(08:30) Error Handling and Dash pipeline integration (09:30) 30 min exercises & 30 min solutions and explanations with Integrating REST api into pipeline (10:30) 1 hour (coding together) tutorial - Adding user selections to your Dash application (11:30) 1 hour Project Work & Questions
Fri	(08:30) Lecture on forecasting and confidence intervals: Drawing conclusions from your datasets. (09:30) 1 hour exercises and solutions in data forecasting (10:30) 1 hour (coding together) tutorial on integrating forecasting into your Dash Applications (11:30) 1 hour Project Work & Questions

Wk. 6	
Mon	(08:30) Feature Engineering for Random Forest (09:30) 30 min exercises in implementing Feature Engineering Pipeline (10:30) 1 hour (coding together) feature engineering on a fresh dataset (11:30) 1 hour project work and questions
Tue	(08:30) Random Forest Deep Dive (Random Forest for continuous variables) (09:30) 30 min exercises in Random Forest (10:30) 1 hour (coding together) tutorial: Visualizing the results of your Random Forest (11:30) Setup a training pipeline
Wed	(08:30) Deep Focus on advanced topics in Statistics (09:30) 30 min exercises & 30 min solutions and explanations & 1 hour (coding together) tutorial on regression, revisited (10:30) 1 hour (coding together) tutorial - integrating regression in your feature engineering pipeline (11:30) 1 hour Project Work & Questions
Thur	(08:30) Advanced Dash Operations (integrating User selections with your Machine Learning algorithm) (09:30) 1 hour exercises on Statistics Re-Cap (10:30) 1 hour (coding together) tutorial on challenge topics in statistics (11:30) 1 hour Project Work & Questions
Fri	(08:30) Technical Interview Drilling (all bootcamp participants practice job interviews + reviewing interviews) (11:30) 1 hour Project Work & Questions

Wk. 7	
Mon	(08:30) Introduction to Large Language Models (09:30) 30 min exercises in Vector Operations (10:30) 1 hour (coding together) Installing and Fine-Tuning a Language Model (11:30) 1 hour project work and questions
Tue	(08:30) Introduction to Retrieval Augmented Generation (RAG) (09:30) 30 min exercises in Vector Search (10:30) 1 hour (coding together) tutorial: Building a Vector Search Database (11:30) 1 hour project work and questions
Wed	(08:30) Introduction to Nearest-Neighbor Search Algorithms + Intro to FAISS (09:30) 30 min exercises & 30 min solutions and explanations (10:30) 1 hour coding time - Nearest Neighbor Search Challenge! (11:30) 1 hour Project Work & Questions
Thur	(08:30) Advanced RAG Principals + Hybrid RAG (09:30) 1 hour exercises on Statistics Re-Cap (10:30) 1 hour coding time - Nearest Neighbor Search Challenge! (11:30) 1 hour Project Work & Questions
Fri	(08:30) Advanced RAG Principals + Hybrid RAG (09:30) 1 hour exercises on Statistics Re-Cap (10:30) 1 hour (coding together) tutorial: Building an advanced hybrid RAG pipeline from scratch (11:30) 1 hour Project Work & Questions

Wk. 8	
Mon	(08:30) Technology Lecture: Docker (09:30) 30 min exercises creating docker containers (10:30) 1 hour (coding together) feature engineering on a fresh dataset (11:30) 1 hour project work and questions
Tue	(08:30) Technology Lecture: Deployment (09:30) 30 min exercises in Deployment of docker containers (10:30) 1 hour (coding together) tutorial: Building and deploying a dash app from scratch (11:30) 1 hour project work
Wed	(08:30) Technology Lecture: Postgres and SQL Alchemy (09:30) 30 min exercises in SQLAlchemy + 30 solution review (10:30) 1 hour (coding together) tutorial: Database query practice (11:30) 1 hour project work
Thur	(08:30) Technology Lecture: BigQuery + Spark + Big Data Methods (09:30) 2 hour technical job-interview question practice (11:30) 1 hour Project Work & Questions
Fri	(08:30) Technical Interview Drilling (all participants practice job interviews + reviewing interviews) (11:30) 1 hour Project Work & Questions