Machine Learning Workshop for Data Scientists

This course is aimed at (aspiring) data scientists, who want to use machine learning in practice, i.e., analyze complex data and build predictive models.

What you will learn:

- What machine learning (ML) is & when you should (not) use it
- Which kinds of problems can be solved with the different ML algorithms
- In what ways insufficient data quality or quantity can derail an ML project
- How the different ML algorithms work; specifically methods for:
  - Dimensionality Reduction
  - Outlier/Anomaly Detection
  - Clustering
  - Supervised Learning (Linear Models, Decision Trees, Ensemble Methods, Kernel Methods)
  - Deep Learning (i.e, Neural Networks)
  - Information Retrieval
  - Recommender Systems
  - Time Series Forecasting
  - Reinforcement Learning
- How to evaluate models and select their hyperparameters
- Common pitfalls & strategies to avoid them:
  - Over- or Underfitting
  - Data & Concept Drift
  - Causality
  - Adversarial Attacks
  - Systematic Bias
  - Explainable ML
- What steps a company can take to successfully use ML in practice
- How to use the central tools of the Python data science ecosystem (Jupyter notebooks, numpy, scipy, matplotlib, pandas, scikit-learn, pytorch/keras)

Prerequisites:

- Good command of the English language (course materials are in English)
- Math skills at university entrance level (especially linear algebra: matrices & vectors, functions)
- Good programming and basic Python skills (e.g., by working through the supplied online tutorial at the beginning of the course (takes approx. 4 hours))
- Have fun puzzling and an interest in new technologies
- Intrinsically motivated to learn new things
- Desirable: First practical experience in analyzing data
- Technical requirements: Python installation on your own computer or access to some server with Jupyter notebooks (e.g., Google Colab, Microsoft Azure workspace, etc.)

Course Procedure:

- Flipped classroom format: Participants study the theory and solve the exercises at their own speed and the results are then discussed in the group sessions
- Duration: 3 weeks with 2 meetings per week (each approx. 2 hours, e.g., Tuesday morning and Thursday afternoon)
- 3 - 6 Participants
- Remote via MS Teams or similar
- All course materials are available online for free

Machine Learning Course Syllabus

The course is separated into 5 parts, that should take about 8 hours each to complete (depending on your Python proficiency; *case study* about 5h). First read the respective sections in the book, then complete the associated exercises. Feel free to schedule an individual coaching session anytime to ask questions or get feedback on your solutions!

Book: https://franziskahorn.de/mlbook/
Exercises: https://github.com/cod3licious/ml_exercises

Part 1:	**Introduction**	[Quiz 1]
	ML with Python	Python tutorial (& installation)
	Data & Preprocessing	[Quiz 2]
	ML Solutions: Overview	[Quiz 3]
Part 2:	**Unsupervised Learning**	
	Dimensionality Reduction	[NB 1] visualize text
	Outlier / Anomaly Detection	
	Clustering	[NB 2] image quantization
	Supervised Learning	
	Different types of models	
	Model Evaluation	[Quiz 4]
Part 3:		[NB 3] supervised comparison:
	Linear Models	- linear models
	Decision Trees	- decision trees
	Ensemble Methods	- random forest
	k-nearest neighbors (kNN)	- kNN
	Kernel Methods	- SVM
	Deep Learning & more	
	Information Retrieval (Similarity Search)	[NB 4] information retrieval
	Deep Learning (Neural Networks)	[NB 5] MNIST with torch (or keras)
	Time Series Forecasting	
	Recommender Systems (Pairwise Data)	
Part 4:	**Avoiding Common Pitfalls**	
		[NB 6] analyze toy dataset
		[NB 7] *predict hard drive failures*
Part 5:	**Reinforcement Learning**	[NB 8] RL gridmove
	Conclusion	[Quiz 5] [Exercise] Your next ML Project