# Machine Learning Course Workbook

– Before the Course –

## Introduction

**ML is everywhere!**
*Where (else) do you use ML in your everyday life incl. work?*

-

– Part 1 –

**Data is the new oil!?**
*What does structured and unstructured data look like?*

- ➔ Structured Data:
- ➔ Unstructured Data:

*Can you think of a decision that you (or someone close to you) made that might have turned out differently if someone had first analyzed some data? Which future decision would you like to make in a data-driven way?*

**What is ML?**
*When are the benefits of ML compared to traditional software?*

*What is the difference between Machine Learning, Artificial Intelligence, and Deep Learning?*

*Take another look at the ML algorithm cheat sheet & try to find an example where you are (or could be) using each of these algorithms either in your everyday life or maybe you even have an idea where one of these algorithms could be used to improve one of your company's products.*

- Dimensionality Reduction:
- Anomaly Detection:
- Clustering:
- Regression:
- Classification:
- Recommender Systems/Information Retrieval:
- Deep Learning:

*What are the benefits of breaking down a complex input-output problem into simpler subproblems?*

*What is the downside of a system composed of multiple ML models?*

**ML history: Why now?**

*What accelerated the rise of ML in the last few years?*

- 
- 
- 

*What is the difference between ANI and AGI?*

**How do machines "learn"?**

*Describe the different learning strategies and what their requirements (in terms of data) are:*

➔ Unsupervised Learning:

➔ Supervised Learning:

➔ Reinforcement Learning:

*What are "features" and what are "labels"?*

➔ Features:

➔ Labels:

*What is the drawback of unsupervised learning methods?*

**Solving problems with ML**

*When should you <u>not</u> use ML?*

*Which kind of ML problems have a high chance of success and when is the outcome uncertain?*

*Which tasks take up most of a Data Scientist's time?*

*What are the two deployment options for an ML model and when should you use which?*

- 
- 

# Data & Preprocessing

**Garbage in, garbage out!**

*What do you think are the most common ways in which datasets in your organization are messy?*

*Which concrete next steps could your organization take to improve their data quality?*

– Part 2 –

## Avoiding Common Pitfalls

*With which stupid baseline should you compare regression and classification models respectively?*

*When is it a really bad idea to evaluate a classification model with the accuracy metric?*

*What does it mean for a model to over- or underfit?*

*Why can a model still be wrong, even though it generates correct predictions for data points from the testset?*

*What are "Adversarial Attacks"?*

*In what ways can a biased model negatively affect users?*

*How can you check whether a model discriminates?*

*What is the difference between data and concept drift?*

*What could be reasons for data or concept drift in your domain / next project?*

## Conclusion

*According to Andrew Ng, what are the 5 steps for a successful AI transformation of a company and where do you think your organization stands in this process?*

1.
2.
3.
4.
5.