

# Machine Learning Kurs Arbeitsbuch

– Vorbereitung –

## Einleitung

### **ML ist überall!**

*Wo (sonst) verwendest du ML in deinem täglichen Leben (inkl. Arbeit)?*

-

– Teil 1 –

### **Daten sind das neue Öl!?**

*Wie sehen strukturierte und unstrukturierte Daten aus?*

- Strukturierte Daten:
- Unstrukturierte Daten:

*Fällt dir eine Entscheidung ein, die du (oder ein Bekannter) getroffen hast, die anders ausgefallen wäre, wenn jemand zuerst Daten analysiert hätte? Welche kommende Entscheidung möchtest du datengetrieben treffen?*

### **Was ist ML?**

*Was sind die Vorteile von ML gegenüber traditioneller Software?*

*Was ist der Unterschied zwischen Machine Learning, Künstlicher Intelligenz und Deep Learning?*

*Schau dir nochmal das [ML Algorithmen Cheatsheet](#) an und versuche für jede Kategorie ein Beispiel zu finden, wo du diese Algorithmen entweder in einer alltäglichen Anwendung verwendest oder, noch besser, wo eure Firma diese Algorithmen einsetzen könnte um ihre Produkte zu verbessern.*

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

*Was für Vorteile hat es, ein komplexes Input-Output Problem in einfachere Teilprobleme zu zerlegen?*

*Was ist der Nachteil an einem System mit mehreren ML Modellen?*

### **ML Geschichte: Warum jetzt?**

*Welche Faktoren beschleunigten den Aufstieg von ML in den letzten Jahren?*

- 
- 
- 

*Was ist der Unterschied zwischen ANI und AGI?*

### **Wie “lernen” Maschinen?**

*Beschreibe die verschiedenen Lernstrategien und ihre Anforderungen (in Bezug auf die Daten):*

- Unsupervised Learning:
- Supervised Learning:
- Reinforcement Learning:

*Was sind “Features” und was sind “Labels”?*

- Features:
- Labels:

*Was ist das Manko von Unsupervised Learning Methoden?*

### **Probleme mit ML lösen**

*Wann sollte man ML nicht verwenden?*

*Welche ML-Probleme haben eine gute Erfolgsaussicht und wann ist der Ausgang ungewiss?*

*Welche Aufgaben beanspruchen die meiste Zeit eines Data Scientists?*

*Was sind die zwei Deployment-Optionen für ML-Modelle und welche Vorteile haben sie?*

- 
-

## Daten & Preprocessing

### **Garbage in, garbage out!**

*Was denkst du sind die häufigsten Probleme, die bei Datensätzen in eurem Unternehmen auftreten?*

*Welche konkreten nächsten Schritte könntet ihr als Organisation unternehmen, um die Datenqualität zu verbessern?*

– Teil 2 –

## Häufige Fehler vermeiden

*Mit welcher dummen Baseline sollte man Regressions- bzw. Klassifikationsmodelle vergleichen?*

*Wann ist es eine ganz schlechte Idee, ein Klassifikationsmodell mit der Accuracy zu evaluieren?*

*Was bedeutet es, wenn ein Modell over- oder underfittet?*

*Warum kann ein Modell trotzdem falsch sein, auch wenn es korrekte Vorhersagen für Datenpunkte aus dem Testset generiert?*

*Was sind “Adversarial Attacks”?*

*Auf welche Weise kann sich ein verzerrtes Modell negativ auf die Nutzer auswirken?*

*Wie kann man überprüfen, ob ein Modell diskriminiert?*

*Was ist der Unterschied zwischen einem Daten und einem Konzept Drift?*

*Aus welchen Gründen könnte es in deinem Bereich / bei deinem nächsten Projekt zum Daten oder Konzept Drift kommen?*

## Fazit

*Was sind laut Andrew Ng die 5 Schritte für eine erfolgreiche KI-Transformation eines Unternehmens und wo denkst du steht euer Unternehmen in diesem KI-Transformationsprozess?*

- 1.
- 2.
- 3.
- 4.
- 5.