

# Machine Learning Kurs Arbeitsbuch

## Einleitung

### **ML ist überall!**

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

-

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

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

-

-

-

*Was ist der Unterschied zwischen ANI und AGI?*

### **Was ist ML?**

*Definiere ML:*

*Was braucht man um ein ML-gestütztes Produkt (also Wert) zu schaffen?*

-

-

-

-

*KI und ML Entwickler, Statistiker und Data Scientists benutzen alle bestimmte Werkzeuge.*

*Was ist der Unterschied zwischen ...*

- ML vs. KI?
- ML vs. Deep Learning?
- ML vs. Statistik?
- ML vs. Data Science?

### **Wie “lernen” Maschinen?**

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

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

## **Wann sollte man ML verwenden?**

*Wodurch kann man mit ML Wert schaffen?*

*Wann sollte man ML nicht verwenden?*

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

## **Probleme mit ML lösen: Workflow**

*Was sind die 3 Hauptschritte um mit ML Wert zu schaffen?*

- 1.
- 2.
- 3.

*Was sollte man überprüfen bevor man ein ML-Projekt startet?*

- 
- 
- 
- 

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

- 
- 

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

## **Daten & Preprocessing**

*Was sind "Features" und was sind "Labels"?*

- Features:
- Labels:

*Wie sehen strukturierte und unstrukturierte Daten aus?*

- Strukturierte Daten:
- Unstrukturierte Daten:

## **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?*

## **ML Lösungen: Überblick**

*Was ist der Output für einen Datenpunkt bei diesen verschiedenen Algorithmus-Kategorien?*

- Dimensionsreduktion:
- Anomalieerkennung:
- Clustering:
- Regression:
- Klassifikation:
- Recommender Systems/Information Retrieval:

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

## 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?*

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

*Wo denkst du steht euer Unternehmen in diesem KI-Transformationsprozess?*