

# Substring - The Data Company

## Machine Learning Projekte starten Tipps und Good Practices

Dataholics#4 Webinar-Reihe



**Nino Müller**  
Partner & Lead Technology  
@ Substring

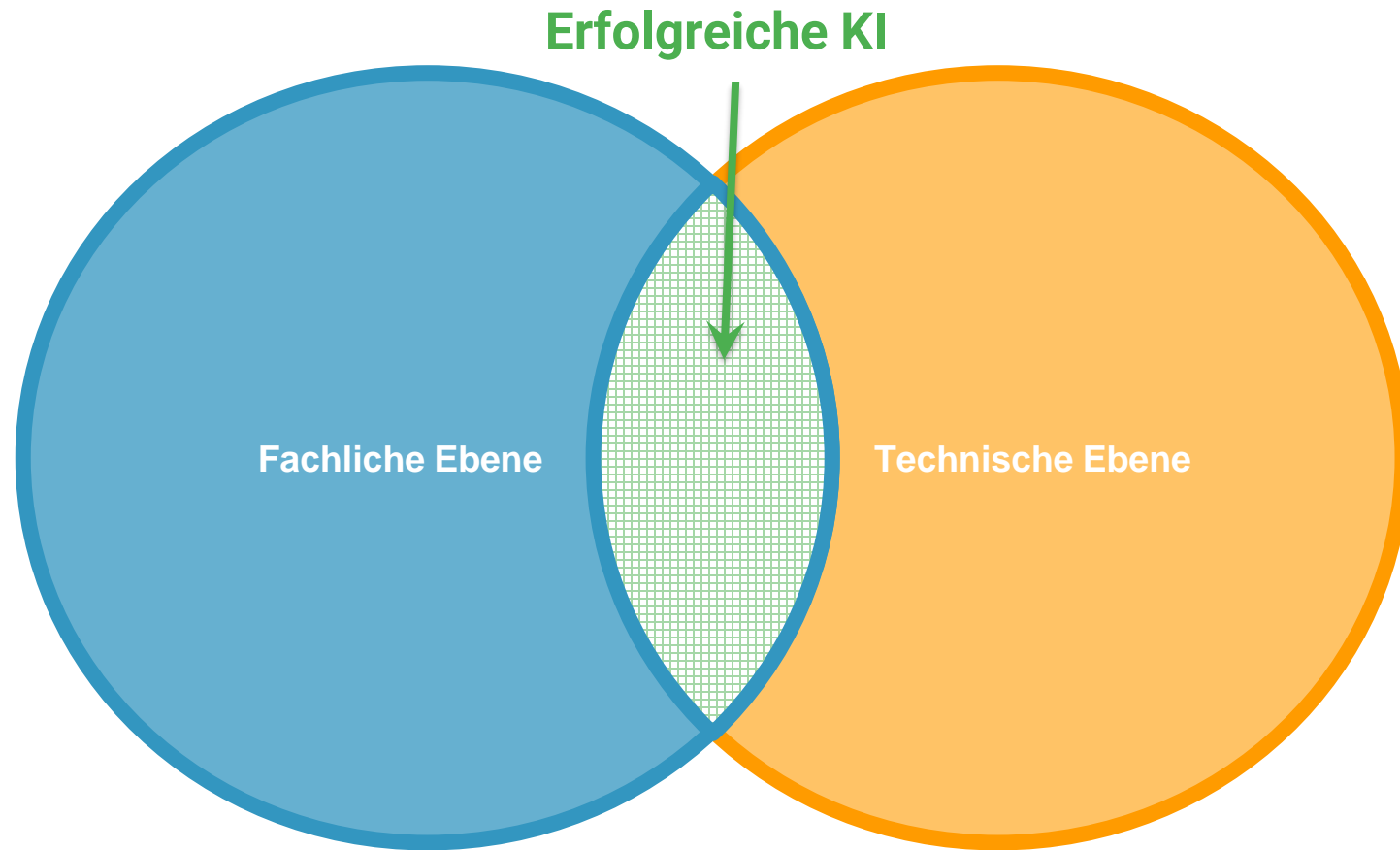
Teilnehmen unter  
**[attend.sl/dtuf](https://attend.sl/dtuf)**



# SO WHAT?

## Was sind konkrete Anwendungsfälle von KI in Unternehmen ?

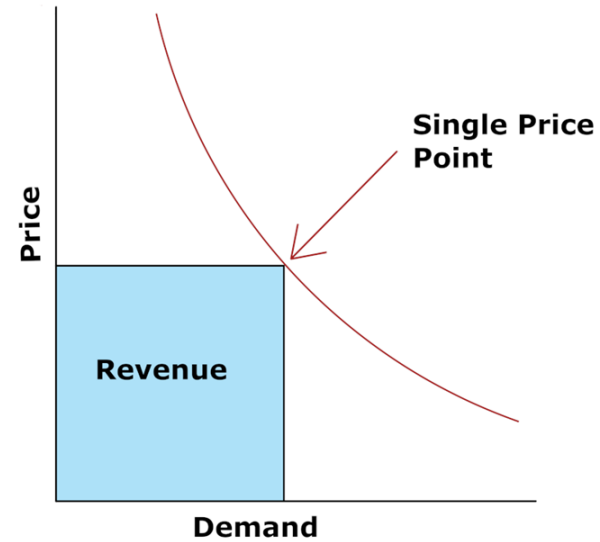
- Verbesserung von Kundeninteraktionen über textbasierte Dialogsysteme
- Spracherkennung/-verarbeitung / Voice Assistents
- Textanalyse (z.B. Spam-Detection)
- Predictive Maintenance
- Erkennung von Abweichungen (Outlier detection)
- Bilderkennung / Klassifikation
- Autonomes Fahren
- Prediction / Scoring



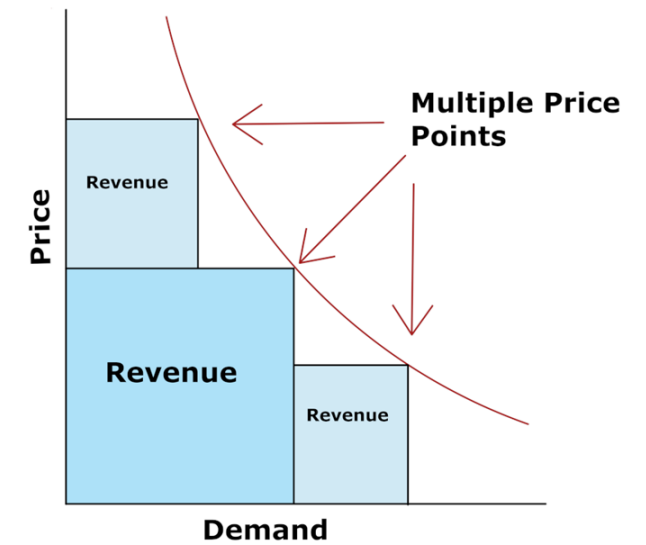
Use Case

# Dynamic Pricing

**Static Pricing**



**Dynamic Pricing**





# Wie geht man das an?



- 1 **Starte** mit dem Business Case
- 2 **Bestandsaufnahme** der Daten
- 3 **Beschreibe die Lösung** um eine klare Definition des Projekts zu erzielen
- 4 **Bestimme** einen Methodenkoffer oder Architektur
- 5 **Data Preprocessing**
- 6 **Modell** erstellen
- 7 **Produkt** erstellen (Operationalisieren)
- 8 **Testen** und anpassen

Vorgehen

# Starte mit dem Business Case



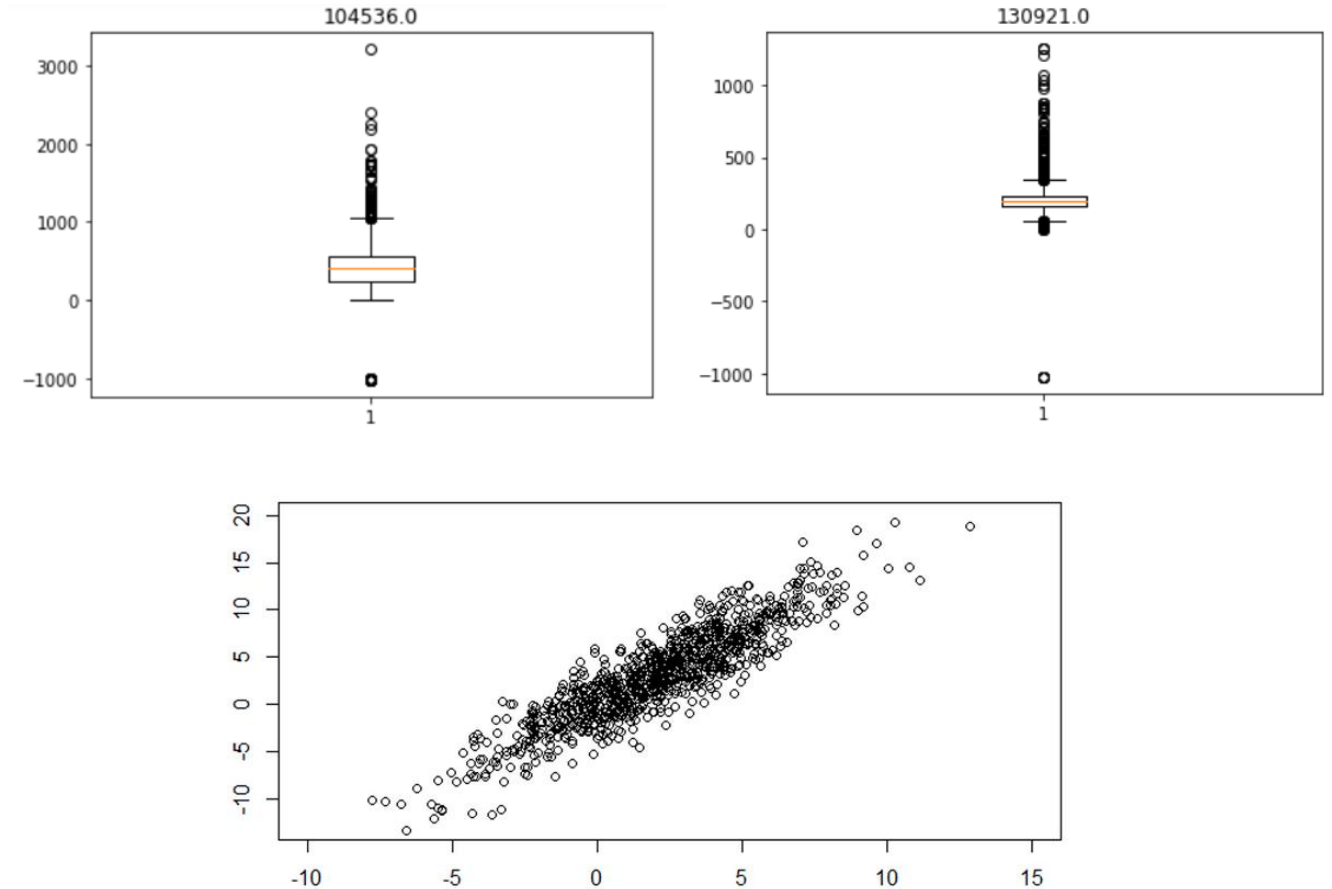
AI Canvas for PoCs			Project Name:
<b>Value</b>	<b>Zielgruppe</b>	<b>Risiken</b>	<b>Problem</b> Hierbei geht es um die Beschreibung des Problems, welches mit Hilfe von KI gelöst werden soll. Im Falle einer Prozessoptimierung wird das bisherige Vorgehen geschildert und begründet, weshalb dieses problematisch ist.
Welcher Mehrwert wird durch die KI generiert (z.B. Effizienz, Umsatzsteigerung oder Kostenreduktion)?	Ein PoC wird in den Regel nicht vollständig ausgerollt, sondern wird in einer klar definierten Gruppe (z.B. Kunden, Mitarbeiter, Produkte) getestet.	Beim Umgang mit grossen Datenmengen und beim Verwenden von KI entstehen Risiken (z.B. Bias in den Daten, Rechtliches, Datenschutz, Transparenz.)	
	<b>Bedingungen</b>	<b>Data Sets</b>	
	Bevor mit dem PoC gestartet werden kann, müssen spezifische Bedingungen erfüllt sein, wie z.B. die Sicherstellung der zeitlichen Ressourcen innerhalb des Teams oder die Verfügbarkeit und Qualität der Daten.	Welche Daten werden für den PoC verwendet? Sind diese Daten bereits vollständig oder müssen sie zuerst analysiert und validiert werden?	
	<b>Ressourcen</b>	<b>Team</b>	
Wie gross ist das Budget? Wie viel Zeit steht für den PoC zur Verfügung? Welche weiteren Ressourcen werden benötigt?	Wer ist Ansprechpartner auf der Business-Seite? Wer ist der Datenexperte? Welche Personen und Abteilungen müssen mit an Bord sein?		
<b>Hypothesen &amp; Metrics</b>			
Die Performance einer KI bzw. der generierte Mehrwert muss zwingend gemessen werden. Dafür sollten klare Hypothesen formuliert und Metriken bestimmt werden, um diese Hypothesen zu bestätigen oder zu verwerfen.			
<b>Lösungsansätze KI PoC</b>			
Wie lautet der KI Use Case? Und handelt es sich dabei z.B. um eine Regressionsanalyse, eine Klassifikation oder ein Clustering?			

Kostenloser  
Download unter



Vorgehen

# Bestandsaufnahme der Daten





Vorgehen

# Lösung & klare Definition des Projekts



Kunde will  
bestellen

 **Modell (KI)**

optimaler Preis

Vorgehen

# Bestimme eine Architektur und einen Prozess



Reproduzierbarkeit

Verantwortlichkeit

Kontinuität

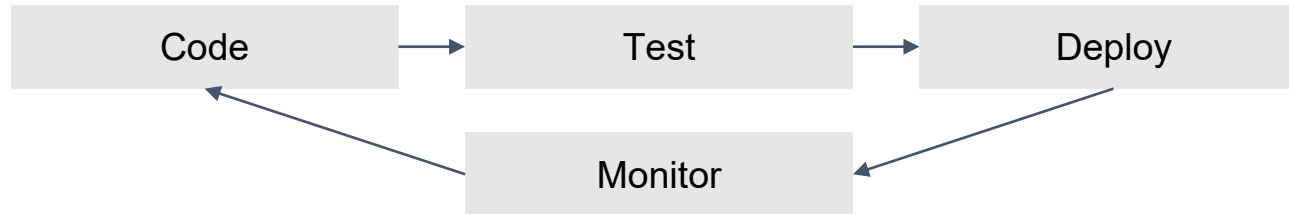
Kollaboration

Vorgehen

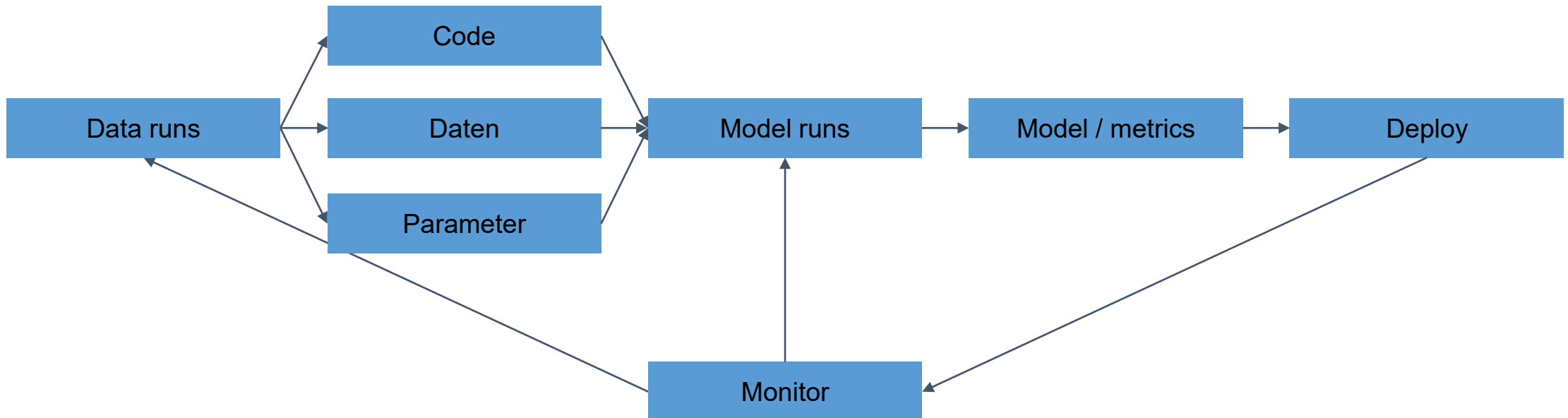
# Bestimme eine Architektur und einen Prozess



# Prozess für Software-Entwicklung



# Prozess für Machine Learning

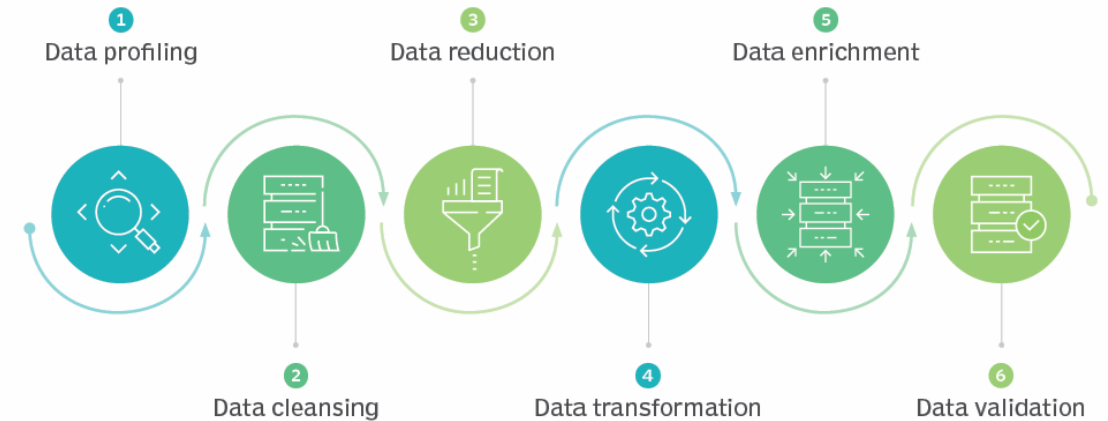


Vorgehen


# Data Preprocessing



## Steps for data preprocessing

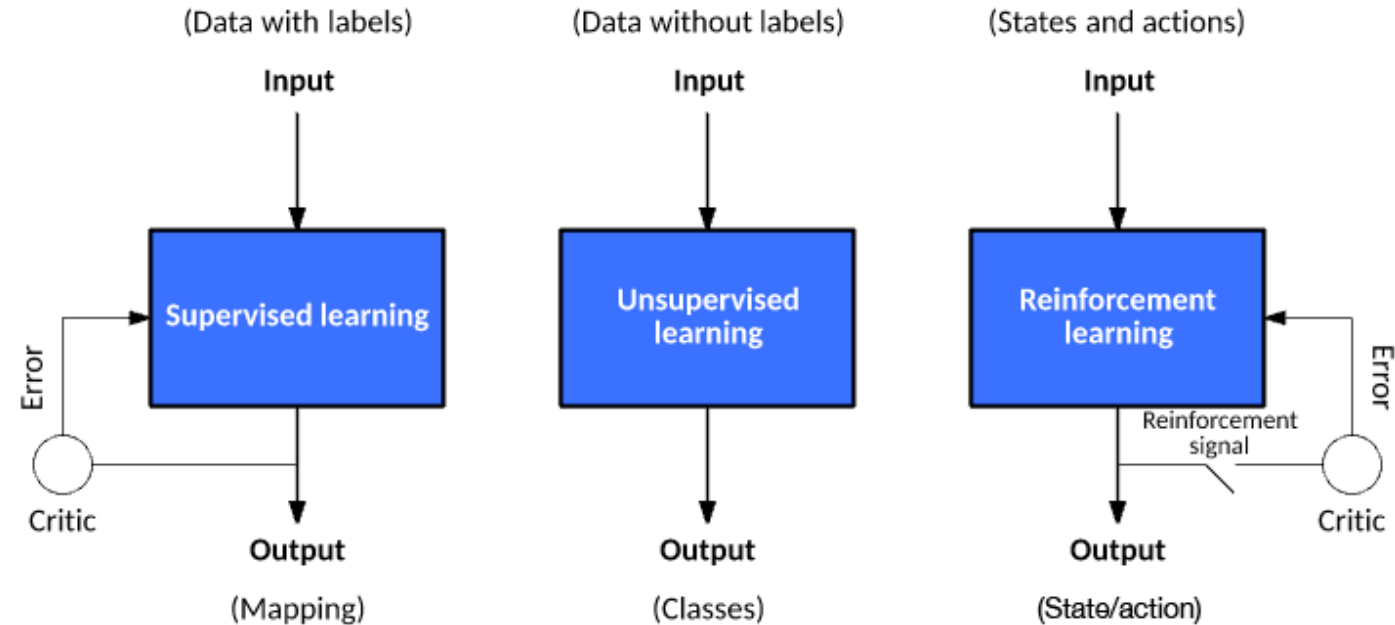


ICONS: ENIS/ANSOV/GETTY IMAGES

©2022 TECHTARGET. ALL RIGHTS RESERVED. 

Vorgehen

# Modell erstellen



Vorgehen

# Produkt erstellen



## Verkaufspreise schwanken – wir sagen weshalb

Unsere Verkaufspreise schwanken, und das teils mehrmals pro Tag. Das sorgt bei der Kundschaft immer wieder für Aufregung. Ein guter Grund, Transparenz zu schaffen und zu zeigen, dass es vor allem externe Faktoren sind, die unsere Verkaufspreise beeinflussen.

Ein klassisches Szenario: Eine Kundin hat sich einen neuen Fernseher gekauft. Zwei Wochen später stellt sie

839.-  
 LG OLED42C24LA (4K, OLED, 2022, 42")  
 ★★★★★

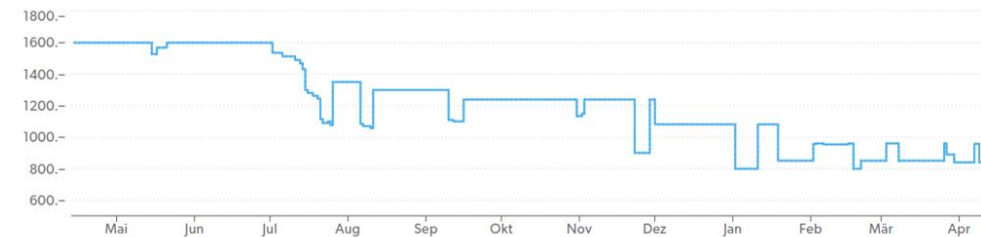
In den Warenkorb

### Verfügbarkeit

### Preisentwicklung

Hier siehst du den jeweils günstigsten Preis des Tages über den Verlauf der Zeit. [Mehr erfahren](#)

3 Monate Alles

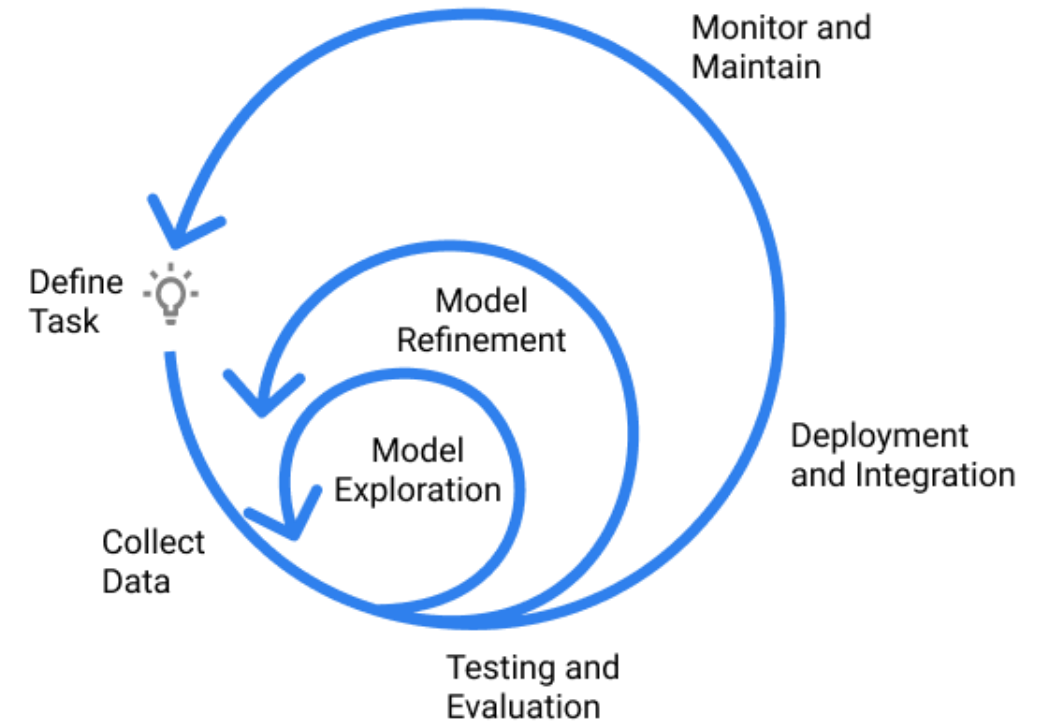


Vorgehen

# Testen und anpassen



## Machine Learning Development Lifecycle



<https://www.jeremyjordan.me/ml-projects-guide/>



# ML-Starter Kit



Substring

# Nächster Dataholics Termin

## Daten nutzbar machen @ Post Immobilien

Wann: Dienstag, 30. Mai 2023

Uhrzeit: 11:15 Uhr - 12:00 Uhr

Anmeldung: <https://share.hsforms.com/1NhZYDsPkRXCh4tB0zJxqfg5mdeq>



Substring GmbH

# The Data Company.

## Services Technical

- Data Science & Machine Learning
- Business Intelligence & Data Discovery
- Data Lake & Data Warehouse
- DataOps & DevOps
- APIs & Microservices für Datenappl.

## Services Business

- Datenlandkarte
- Daten Strategien
- Business Analysen
- Datenarchitektur Design
- Schulungen zu Datenkompetenzen

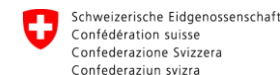
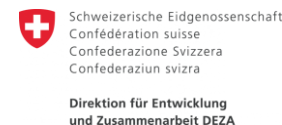
## Lehraufträge & Weiterbildung

- CAS Digitale Organisation (BFH)
- CAS Data-Driven Organization (BFH)
- CAS New Quality Management (SAQ/HWZ)

## Team

- 10 Mitarbeitende mit Leidenschaft für Daten
- Interdisziplinäres Team (Didaktik, Ökonomie, Statistik, SW-Entwicklung)
- + 20 Jahre Erfahrung in der Entwicklung von digitalen Data Lösungen

## Zufriedene Kunden



HAUTE ÉCOLE DE GESTION  
HOCHSCHULE FÜR WIRTSCHAFT  
SCHOOL OF MANAGEMENT

Fribourg  
Freiburg





# Substring - The Data Company

---

*This wasn't very data-driven of you!*