

**Online-Zertifikatslehrgang** 

### Data Analyst IHK

Die neue Generation digitaler IHK-Weiterbildungen



IHK■Die Weiterbildung

### Modul 4 - Datenprojekte – bewerten, planen und umsetzen

#### > Grundlagen des Agilen Arbeitens

#### > Einleitung und Ist-Analyse:

- Beschreibung der Ausgangssituation und des Hintergrundes der Daten
- · Erläuterung der Zusammensetzung der Daten,
- · Bewertung der Datenqualität.

#### > Backlog und Projektziele

- Anforderungen: User Storys und Potenziale.
- Bewertung von Aufwand und Ertrag der Elemente im Backlog.
- Priorisierung der Backlogelemente und Auswahl der erreichbaren Projektziele

#### > Hypothesen:

- Überprüfung der Machbarkeit der Funktionen für die User Storys und Potenziale
- Zuordnung der Hypothesen zu User Storys und Potenzialen

#### Metriken:

- Wie lassen sich Fortschritt und Erfolg des Projektes und die Qualität der Analyse bzw. des Dashboards im Einsatz messen?
- Definitions of done: Akzeptanzkriterien für das Erreichen der ausgewählten Ziele

#### Machine Learning Canvas:

• Funktionale und technische Projektübersicht des ersten Sprints

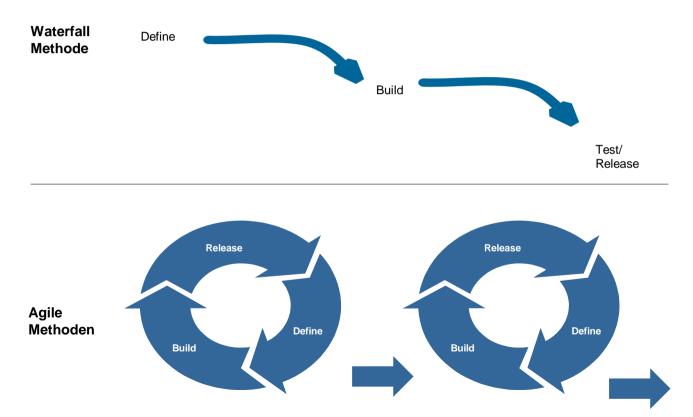
#### > Planung des ersten Sprints:

• Formulierung der einzelnen Arbeitsschritte und Aufgaben sowie des dafür benötigten Aufwandes.





# Agile Methoden verändern klassische Entwicklungsprozesse.



# Agile Methoden für Datenprojekte Werte und Prinzipien

### Agile Methoden

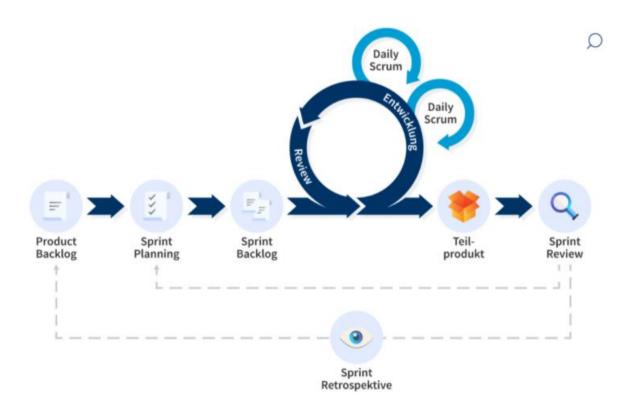
#### Werte

- Menschen und ihre Interaktionen sind wichtiger als Prozesse und Werkzeuge
- Funktionierende Software ist wichtiger als vollständige und ausführliche Dokumentation
- ...

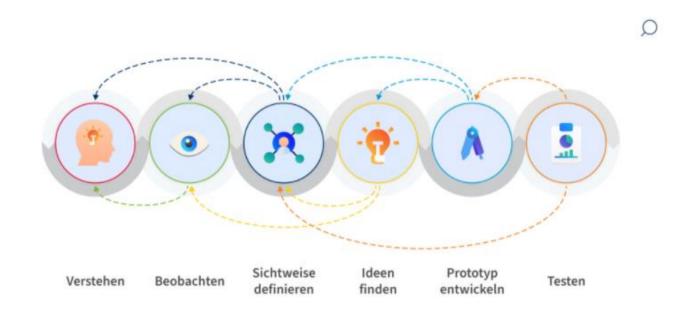
#### **Prinzipien**

- Kundenzufriedenheit durch frühe und kontinuierliche Lieferung von Software
- Change Requests werden als Normalität betrachtet (auch in späten Entwicklungsstadien) und sind willkommen
- Kommunikation von Angesicht zu Angesicht ist die beste Form der Kommunikation
- ..

### Arbeiten mit SCRUM

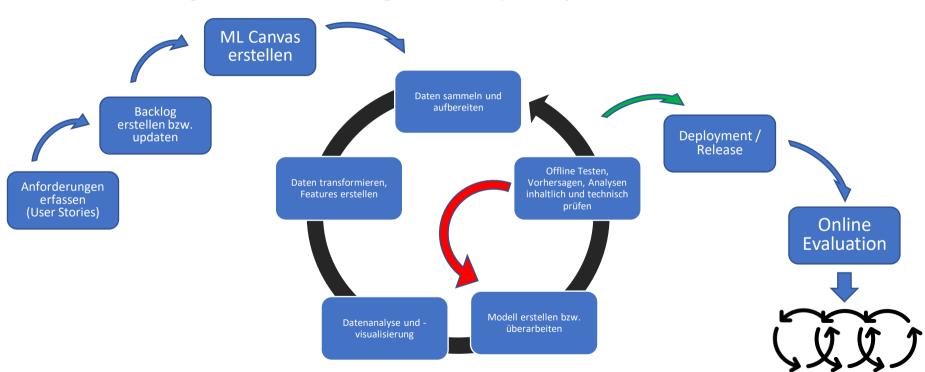


### Design Thinking



### Machine Learning Cycle

Ein iteratives Vorgehen macht den Kern agiler Data Analytics Projekte aus.



### Die Vorbereitungsphase

Anforderungen erfassen (User Stories)



Backlog erstellen bzw. updaten



ML Canvas erstellen Typische Fragen zu Projektbeginn sind:

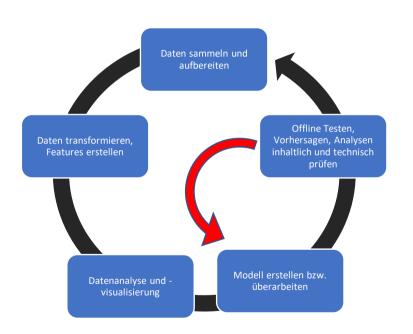
- Welcher Nutzen oder Mehrwert kann geschaffen werden?
- Wer ist der Anwender bzw. die Zielperson, für die der Mehrwert entstehen soll?

Wurden alle Anforderungen gesammelt und priorisiert, wird daraus im nächsten Schritt ein Entwicklungs-Backlog.

Wichtige Aspekte: Priorisierung, Detaillierung, Aufwandsschätzung, Bewertung

Um den Überblick in einem neuen Projekt zu behalten, lohnt es sich, am Projektbeginn einen Machine Learning Canvas auszufüllen.

### Der Entwicklungszyklus



#### 1. Daten sammeln und aufbereiten

- Daten sammeln
- Daten verstehen (Explorative Datenanalyse)
- Qualitätsprobleme korrigieren

#### 2. Daten transformieren, Features erstellen

#### 3. Datenanalyse und Visualisierung

- Kennzahlen evaluieren
- Geeignete Algorithmen bestimmen
- Baseline, an der sich das Modell misst

#### 4. Modell erstellen bzw. überarbeiten

- · Modell mit Trainingsdaten aufbauen
- · Gewählte Algorithmen gegenüberstellen

#### 5. Offline testen

- Inhaltliche Prüfung: Bezug, Genauigkeit, Over- oder Underfitting, Beeinflussung der Daten
- Technische Prüfung: Stabilität, Aufwand, Verfügbarkeit

### Die Umsetzung

Deployment / Release Kommt das Modell zum Einsatz, erfolgt die Projektübergabe einschließlich der gesamten Dokumentation. Der Auftraggeber untersucht,

- ob das System seine Anforderungen erfüllt und
- ob die Fragen mit akzeptabler Genauigkeit beantwortet werden.

Online Evaluation Es ist wichtig, das Modell kontinuierlich auf seine Funktionalität zu prüfen:

- Wie verhält sich das Modell bei Änderungen der Daten?
- Wie sind die Daten zusammengesetzt? Ist das Modell noch repräsentativ?
- Sind Aussage und Bedeutung der Daten noch gleich oder haben Variablen inzwischen eine andere Bedeutung?





## Weshalb eine Situations- bzw. Ist-Analyse durchführen?

- Datenprojekte starten selten auf der grünen Wiese. Oft gibt es in Betrieben bereits Daten- und Datenverarbeitungsstrukturen.
- Verschiedene Prozesse greifen bereits ineinander und es bestehen häufig Abhängigkeiten zu Datenverarbeitungen und -analysen.
- Wichtig ist am Beginn eines Datenprojekts, Prozesse und Ressourcen, die es bereits gibt, zu identifizieren.
- Dann kann betrachtet werden, wo die Arbeit mit Daten vertieft werden kann und wo bestehende Strukturen vorsichtig gewahrt werden müssen.

### Was ist eine Ist-Analyse?

Bei einer Ist-Analyse werden bestehende Prozesse und Strukturen beleuchtet, um den Ist-Zustand der Arbeitsabläufe im Unternehmen oder in einzelnen Abteilungen zu erkennen und zu beschreiben.

Mit der Beschreibung des Ist-Zustandes wird definiert,

- welche Ressourcen (mittels Ressourcenanalyse) und Prozesse (mittels Prozessanalyse) für das Datenprojekt relevant sind,
- > auf welche zurückgegriffen und welche verändert werden können und
- welche Funktionen essenziell für andere Nutzer/Prozesse sind und daher erhalten bleiben müssen.

### Analyse-Typen

Mit einer Ressourcenanalyse wird übersichtlich der komplette Umfang der Ressourcen dargestellt, die dem Unternehmen bereits zur Verfügung stehen. Dazu zählen Daten und Datenquellen, IT-Strukturen (Datenbanken, Clouds, Schnittstellen etc.) und Software-Lösungen für das Erstellen, Extrahieren und Bearbeiten von Daten.

Mit der Prozessanalyse wird beschrieben, wie die Ressourcen im Unternehmen angewendet werden. Hier werden interne Abläufe sowohl strukturell als auch aus technischer Sicht beleuchtet und bewertet, welche Abläufe genutzt und beibehalten werden können und welche angepasst bzw. optimiert werden sollten. Das Ergebnis ist eine Prozessdokumentation, die auch als Basis zur Ableitung neuer Potenziale oder Verbesserung bestehender Geschäftsabläufe verwendet werden kann.

### Mapping

Die Komplexität von Projekten (und damit auch von Datenprojekten) steigt mit der Vielfalt der enthaltenen Elemente, ihren Beziehungen untereinander und den Zielen des Projekts. Wie soll man das überschaubar halten?

Mit Mapping oder auch Mindmapping kann man Begriffe, Konzepte, Ideen oder Erfahrungen zu einem Thema zu strukturieren und visualisieren. Dabei entstehen abstrahierte Karten in Form von räumlichen Darstellungen, die die betrachteten Elemente in zueinander Beziehung setzen.

Diese Methode ist sehr hilfreich, um sich einerseits einen schnellen Überblick und Verständnis zu einem Thema zu verschaffen, und anderseits, um komplexe Sachverhalte zu entwirren und ihre Wechselwirkungen verständlich darzustellen.

Je nach Zielsetzung und Anwendungsbereich gibt es verschiedene Formen von Mapping, z. B.

- Konzeptmapping
- Prozessmapping
- Strategiemapping

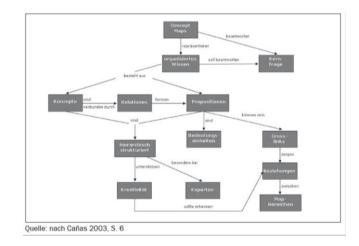
### Konzeptmapping

Mit Konzeptmapping entsteht eine Art "Wissenslandkarte", die man auch Konzeptkarte (englisch Concept Map) nennt.

Eine Konzeptkarte ermöglicht einen guten Überblick über die Zusammenhänge innerhalb eines komplexen Sachverhalts.

Zur Methode des Konzeptmappings gehören:

- Erkunden von Begriffen, die ein Konzept repräsentieren können
- Erfassen gefundener Begriffe als einzelne Themenboxen
- Anordnen der Themenboxen (virtuell oder auch real zum Beispiel auf einem Whiteboard)
- Verbinden der Elemente mit Linien (ungerichtete Beziehungen) und Pfeilen (gerichtete Beziehungen)
- Beschriften der Linien und Pfeile (zum Beispiel "ist Teil von"; "wird erzeugt durch"; "führt zu").
- Veranschaulichen der Beziehung durch Symbole (optional)

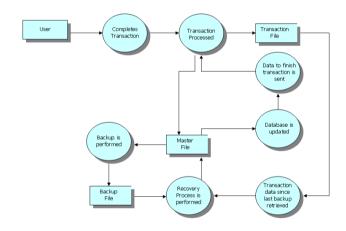


https://www.die-bonn.de/zeitschrift/12011/lehrlernforschung-01.pdf

### Prozessmapping

Prozessmapping ist eine effektive Methode zur Veranschaulichung komplexer Wechselwirkungen in Prozessen. Zur Methode des Prozessmappings gehören:

- Darstellen von Prozessabläufen (häufig als Flussdiagramm)
- Analysieren der einzelnen Prozessschritte
- Zerlegen des Prozesses in kleinste, nicht mehr sinnvoll teilbare Handlungen
- Neugestalten des Prozesses unter Berücksichtigung der Effizienz



https://de.wiktionary.org/wiki/Datenflussdiagramm #/media/Datei:Backup-DFD.png



#### **BREAKOUT-SESSION**

Gruppenarbeit zum Thema "Ist-Analyse" danach weiter im Plenum



Machen Sie sich mit den Daten und der Ausgangssituation vertraut:

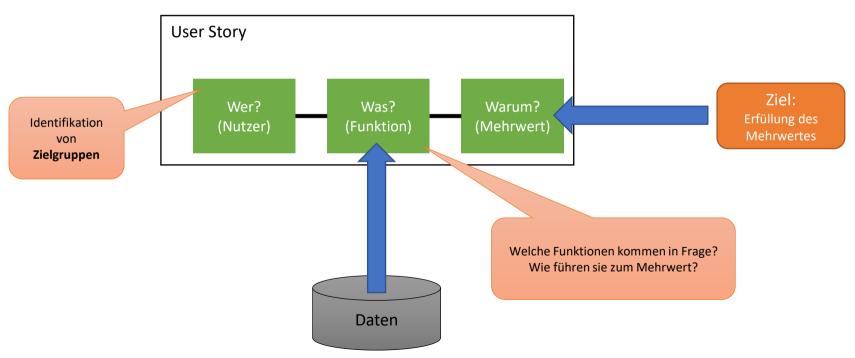
- Beschreiben Sie die Situation aus der die Daten stammen.
- Welche Daten stehen zur Verfügung und was sagen sie aus?
- Wie steht es um Verwendbarkeit und Qualität der Daten?





Wie lassen sich Anforderungen für Projekte erfassen?

### 1. User Storys



### Mehrwert formulieren mit User Storys

Die Basis eines Datenprojektes ist der Mehrwert, der für eine Zielperson oder Zielgruppe geschaffen werden soll.

Zur Erfassung eignet sich die Verwendung von **User Storys**. Sie formulieren die Motivation für ein Datenprojekt aus der Sicht einer Person, die später einen Vorteil aus dem Datenprojekt ziehen wird.

Eine User Story ist so aufgebaut:

"Wer möchte was und warum?"

Rolle (Wer?)	Wer fordert etwas an? Der Anforderer ist meist der spätere Nutzer des Systems oder der Nutznießer der Lösung.
Funktion (Was?)	Was wünscht sich der Anforderer? Je klarer und präziser die Anforderung in der User Story formuliert wird, desto hilfreicher ist sie für die Realisierung.
Nutzen (Warum?)	Warum wird eine Lösung gebraucht? Die Begründung für ein Datenprojekt liefert weitere Informationen zur richtigen Umsetzung.

### Beispiele für User Storys

"Wer möchte was und warum?"

- Der Geschäftsführer möchte einen monatlichen Statusbericht der Abteilungen, um faktenbasierte Entscheidungen treffen zu können.
- Der Einkäufer möchte eine Voraussage der benötigten Produktionsmittel für das kommende Halbjahr, um Einkäufe besser planen und verhandeln zu können.
- Der Marketingleiter möchte eine Klassifizierung von Kundentypen, um so Kunden individueller ansprechen zu können.

#### 2. Potenziale sehen

Potenziale können aus Ist-Analysen und User-Storys abgeleitet werden. Die erkannten Potenziale bilden die Basis eines Datenprojekts.

Potenziale betreffen häufig die Datenqualität.

- Unzuverlässige und fehlerhafte Quellen und Verarbeitungssysteme werden identifiziert und können durch prozesssichere Ressourcen ersetzt werden.
- Durch Automatisierung der Datenextraktion und -verarbeitung k\u00f6nnen manuelle Arbeitsschritte minimiert und somit Prozesszeit eingespart und Fehler vermieden werden.
- etc.



#### **BREAKOUT-SESSION**

Gruppenarbeit zum Thema "Projketziele" danach weiter im Plenum



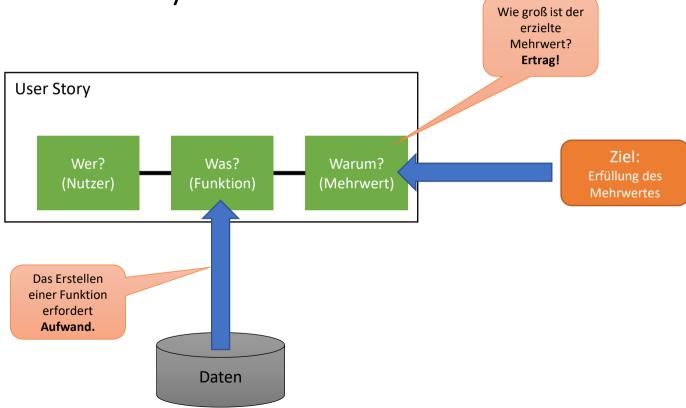
- 1. Definieren Sie anhand von User Storys Mehrwerte für das Datenprojekt.
- 2. Welche Potenziale lassen sich aus der Ist-Analyse ableiten?





Wie lassen sich Projektziele aus dem Backlog auswählen?

### Ziele – User Storys



### Priorisierung von Projektzielen

Schritt 1: Ermittlung der Low Hanging Fruits

Schritt 2: Einsatz des Pareto-Prinzips

Schritt 3: "Minimum Viable Product" als Synthese aus den Low Hanging Fruits und dem Pareto-Prinzip

### Low Hanging Fruits

Low Hanging Fruits ("tief hängende Früchte") bezeichnen Aufgaben und Ziele, die einfach erledigt und erreicht werden können und dabei trotzdem einen möglichst großen Ertrag bringen. In Projekten oder Geschäftsplänen werden die Low Hanging Fruits immer als erstes bearbeitet.

Die Low Hanging Fruit Matrix bietet Anhaltspunkte und eine Methode, um Aufgaben sinnvoll zu kategorisieren und priorisieren.

#### Low Hanging Fruits:

- wenig Aufwand, hoher Ertrag
  - → als erstes bearbeiten und Erfolge ernten

#### Cream of the Crop:

- · viel Aufwand, hoher Ertrag
  - → kontinuierlich weiterbearbeiten und Synergieeffekte nutzen

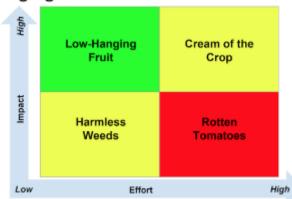
#### Harmless Weeds:

- wenig Aufwand, wenig Ertrag
  - → bei Gelegenheit mitnehmen, aber sehr niedrige Priorität

#### **Rotten Tomatoes:**

- viel Aufwand, wenig Ertrag
  - → nach Möglichkeit ersetzen oder beenden

#### **Low-Hanging Fruit Matrix**

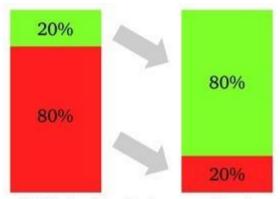


### Das Pareto-Prinzip

Das Pareto-Prinzip beruht auf der Beobachtung von Vilfredo Pareto, dass viele Verteilungen exponentiell verlaufen und somit 80 % der Werte auf 20 % der Teilelemente bzw. Ereignisse fallen.

Auch das Pareto-Prinzip hilft, Aufgaben oder Ziele zu priorisieren. Es nimmt an, dass man in einem Projekt mit einem Einsatz von nur 20 % bereits ein 80 %-iges Ergebnis erreicht. Umgekehrt erfordert das Erreichen der übrigen 20 % einen Einsatz von 80 %.

Daraus lässt sich eine Regel ableiten: Für einen guten Erfolgswert ist häufig zunächst nur ein geringer Aufwand notwendig. Für ein sehr gutes oder perfektes Ergebnis wird der Aufwand jedoch unverhältnismäßig höher.



80% der Ergebnisse werden in 20% der Gesamtzeit erreicht.

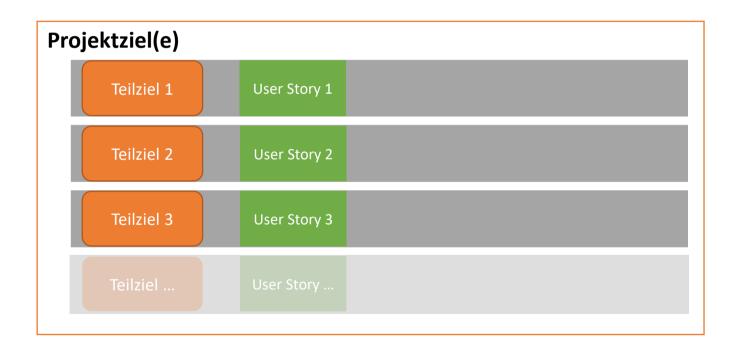
### Synthese: "Minimum Viable Product"

Bei der Entwicklung von Produkten und Dienstleistungen ist es häufig sinnvoll, das geplante finale Endprodukt in "leicht" erreichbaren benutzbaren Zwischenstände zu erzeugen, die aber dem Nutzer bereits einen Mehrwert bieten.

Mit diesen sogenannten "minimal viable products", also minimal einsetzbaren Produkten, erzielt man mit einem Teilaufwand nach dem Pareto-Prinzip bereits Erfolge und erntet die Low hanging fruits mit ersten Zielerfüllungen.

Es muss dabei stets auf die Zufriedenheit des Nutzers geachtet werden. Ist der erbrachte Mehrwert nur mit spürbaren Einschränkungen zu nutzen, kann schnell der Eindruck entstehen, dass es sich um ein "Bananenprodukt" handelt, also dass es erst beim Nutzer reift. Der dadurch entstandene Imageschaden kann die Bereitwilligkeit der Nutzer für die Verwendung der weiterentwickelten Produkte deutlich senken.

## Projektziele setzen sich aus mehreren Teilzielen zusammen





#### **BREAKOUT-SESSION**

Gruppenarbeit zum Thema "Priorisierung von Projektzielen" danach weiter im Plenum



- Bewerten Sie Ihr Backlogelemente nach Aufwand und Ertrag
- Welche Ziele wären Low hanging fuits?
- Wie könnte ein erstes MVP aussehen?

- Die Teilnehmer werden nach ihren Projektgruppen auf Räume verteilt.
- Legen Sie innerhalb der Arbeitsgruppen fest, wer dokumentiert und wer im Anschluss die Ergebnisse vorträgt.





# Experimenteller Charakter von Datenprojekten

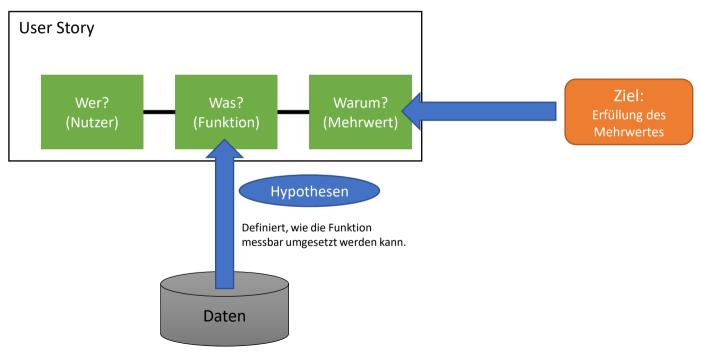
Datenprojekte haben zu Beginn häufig einen experimentellen Charakter. Es gibt Vorstellungen und Ideen, aber eine klare Umsetzungsvorstellung und das Wissen, ob ein Projekt überhaupt machbar ist und zu einem Ergebnis führt, fehlen.

Insbesondere wenn ein größerer Aufwand und hohe Investitionen für die Umsetzung des Projekts notwendig sind, sollte geklärt werden, ob die zu erwartenden Ergebnisse den Einsatz rechtfertigen.

Es empfiehlt sich deshalb, Etappenziele mit Aufgaben, Messwerten und Konsequenzen zu formulieren, um sie zu überprüfen. So kann verhindert werden, dass man sich in Datenprojekten verläuft und seine Ziele verfehlt.

# Wie lässt sich die Machbarkeit von Projektzielen überprüfen?

# Ziele – User Storys - Hypothesen



# Hintergründe und Zusammenhänge verstehen

Beim Aufstellen von Hypothesen sollte darauf geachtet werden, möglichst eindeutige Zusammenhänge zu formulieren und sie mit klaren Messwerten zu verknüpfen.

Dadurch lassen sich Ziele beschreiben, Hindernisse erkennen und Auswirkungen auf die Analyse abschätzen.

Es gibt zwei Möglichkeiten, die Sachlage besser zu verstehen:

- Datensichtung mit einer explorativen Datenanalyse (EDA)
- 2. Expertenbefragung: Bei komplexeren Sachlagen hilft es zunächst, die verschiedenen Abhängigkeiten mithilfe von Mappings zu visualisieren und voneinander zu trennen.

# Entwickeln von Hypothesen

Zum experimentellen Charakter gehören Hypothesen. Um Hypothesen aufzustellen, geht man vom Groben ins Feine und vom Allgemeinen ins Konkrete:

1. Man beginnt zunächst mit einer Annahme. Annahmen sind sehr allgemeine und weiche Aussagen. Sie geben eine grobe Richtung vor.

"Personen, die sich für das Thema Ökostrom interessieren, bewegen auch Themen wie Nachhaltigkeit und Klimawandel."

2. Der nächste Schritt ist das Formulieren einer These. Thesen sind konkreter und stellen einen Zusammenhang zwischen Zuständen oder Objekten her, der grundsätzlich überprüft werden kann. Eine These enthält allerdings noch keine messbare Dimension.

"Personen, die sich mit dem Klimawandel beschäftigen, beziehen häufig Ökostrom."

3. Mit der Messbarkeit des Zusammenhangs wird aus einer These eine Hypothese.

"Die Wahrscheinlichkeit, dass Personen, die sich auf Social Media zum Thema Klimawandel austauschen, auf Ökostrom umsteigen, ist größer als bei Personen, die sich bisher nicht zum Thema Klimawandel auf Social Media geäußert haben."

# Projektstatus bewerten

Hypothesen können dabei helfen, den Projektstatus zu beurteilen:

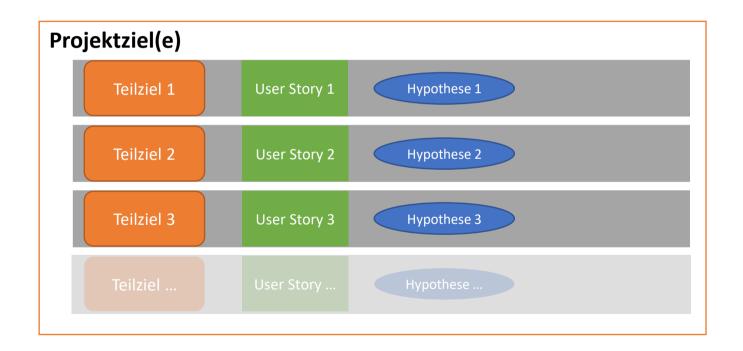
Projektziel: Potenzielle Kunden für das neue Ökostromangebots identifizieren und die Kundenansprache durch Social Media Monitoring optimieren.

Hypothese 1: "Die Wahrscheinlichkeit, dass Personen, die sich auf Social Media zum Thema Klimawandel austauschen, auf Ökostrom umsteigen, ist größer als bei Personen, die sich bisher nicht zum Thema Klimawandel auf Social Media geäußert haben."

Hypothese 2: "Werden weitere Interessen der potenziellen Kunden im Zusammenhang mit dem Ökostromangebot berücksichtigt (zum Beispiel gesunder Lebensstil und Zero Waste), steigt die Wahrscheinlichkeit, dass sie auf unser Angebot umsteigen, zusätzlich."

Je nach Komplexität eines Projekts können mehrere Hypothesen parallel oder nacheinander verfolgt werden. Wird das Datenprojekt in einem agilen Prozess bearbeitet, sollte nicht zu tief und langfristig geplant werden.

# Projektziele setzen sich aus mehreren Teilzielen zusammen





## **BREAKOUT-SESSION**

Gruppenarbeit zum Thema "Hypothesen und Mapping" danach weiter im Plenum



### Stellen Sie Hypothesen zu den Daten auf:

- Welche Zusammenhänge bestehen in den Daten?
- Lässt sich die Erreichung von Funktionen und Mehrwerten durch Hypothesen verifizieren?
- Formulieren Sie einige Hypothesen zu den erstellten User Stories.

- Die Teilnehmer werden nach ihren Projektgruppen auf Räume verteilt.
- Legen Sie innerhalb der Arbeitsgruppen fest, wer dokumentiert und wer im Anschluss die Ergebnisse vorträgt.





Wie lassen Fortschritt des Projektes bzw. das Erreichen der Ziele messen?

# Projektmetriken

Metriken zur Bewertung von Projekten lassen sich in 3 Kategorien einteilen:

#### 1. Rahmenmetriken

Rahmenmetriken kommen häufig aus dem klassischen Projektmanagement und sind Messwerte für grundsätzliche Eigenschaften und Ziele von Projekten.

#### 2. Produkt- und Qualitätsmetriken

Bei Produkt- und Qualitätsmetriken geht es direkt um das Produkt oder die Dienstleistung, die entwickelt bzw. umgesetzt werden sollen, und in welcher Qualität sie vorliegen.

#### 3. Teammetriken

Teammetriken umfassen die Leistung, Entwicklung und Qualifikation von Teams und Teammitgliedern.

# Rahmenmetriken

## 1. Budget und Ressourcen

Hier werden Vorgaben und Grenzen der zur Verfügung stehenden Mittel und Ressourcen im Projekt gemessen und bewertet.

Budgets und Ressourcen müssen kontinuierlich überwacht werden, um frühzeitig Überschreitungen und Gefährdung der Projektversorgung zu erkennen.

Typische Leistungskennzahlen, mit denen Budget und Ressourcen eines Projekts überwacht werden können:

- Planwert
- Istkosten
- Fertigstellungswert
- Kosten-Performance-Index: Fertigstellungswert (Earned Value, EV) / Istkosten (Actual Cost, AC)

# Rahmenmetriken

### 2. Zeit

Die Zeitleiste eines Projekts umfasst in erster Linie das Abschlussdatum (beziehungsweise den Termin für die Produkteinführung), aber auch wichtige Meilensteine und Aufgaben-Deadlines, die zur Erfüllung der Projektziele erforderlich sind.

Zu den Leistungskennzahlen, mit denen Projektfortschritt gemessen wird gehören:

- Prozentsatz der abgeschlossenen Aufgaben
- Überfällige Projektaufgaben
- Verpasste Meilensteine
- Performance-Index f
  ür Zeitpl
  äne: Fertigstellungswert (Earned Value, EV) / Planwert (Planned Value, PV)

# Rahmenmetriken

## 3. Erfolg und Erreichen von Geschäftszielen

Hierbei geht es darum, wie gut ein Projekt im Vergleich zum Business Case abschneidet:

- Hat es die erwarteten Vorteile geschaffen (materiell und immateriell)?
- Wurde die erhoffte Rendite erzielt (ROI)?
- Wurden mehr Kunden erreicht oder die Kundentreue erhöht?

# Produkt- und Qualitätsmetriken

## Mehrwert und Werteversprechen

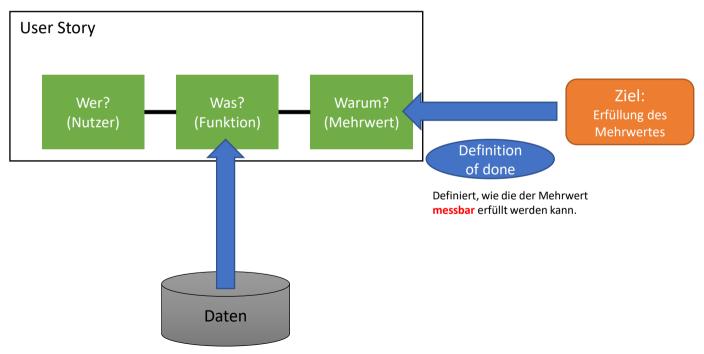
Beim Entwickeln der Produkt- und Dienstleistungsmerkmale stehen bestimmte Eigenschaften im Vordergrund, die beispielsweise durch User Storys oder durch die Methode des Value-Propositon ermittelt wurden.

Werden diese Eigenschaften und Merkmale als qualitätsentscheidend definiert, sollte auch gleich eine (möglichst quantitative) Bewertung der Erfüllung bzw. des Erfüllungsgrades definiert werden (definition of done).

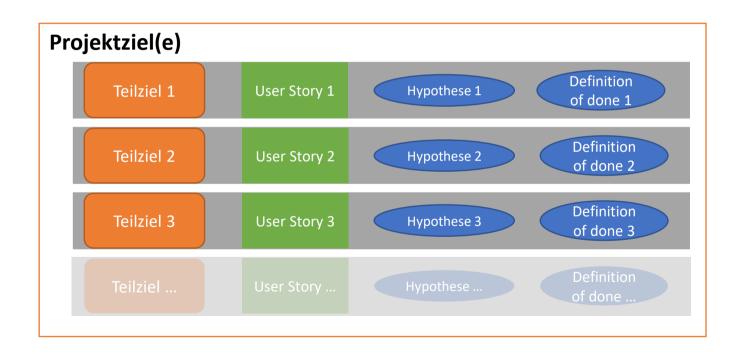
Im agilen PM steht das Erreichung des Mehrwertes bzw. des Werteversprechens im Vordergrund und nicht so sehr, wie dies erreicht wird.

Dies ist wichtig, wenn man den Erfolg messen will. So muss die nötige Flexibilität gegenüber dem "Wie?" (Umsetzungsweg) gewahrt bleiben, während man das "was?" (Mehrwert) beurteilt.

# Ziele – User Storys – Definition of done



# Projektziele setzen sich aus mehreren Teilzielen zusammen



# Produkt- und Qualitätsmetriken

### Kundenzufriedenheit

Gerade bei agilen Projekten ist dies eines der wichtigsten Messkriterien. Produkte und Dienstleistungen werden iterativ erstellt, d.h. es werden bereits Release-fähige Teillösungen (MVP, minimum viable products) an den Kunden geliefert, während das Produkt im nächsten Teilschritt weiterentwickelt wird.

Die Kundenzufriedenheit bietet wichtige Informationen für die Weiterentwicklung des Produktes bzw. der Dienstleistung und kann einem Projekt wichtige Informationen zur Qualität der Ergebnisse, dem allgemeinen Kundenerlebnis sowie zum Kundenservice geben.

Es gibt mehrere Methoden, um Feedback von Kunden zu erhalten und zu analysieren. Dazu gehören:

- Social Media Listening
- Kundenumfragen (einschließlich offener Fragen mit freier Antwortmöglichkeit, nicht nur Bewertungen nach Zahlen)

# Team-Metriken

Das Projektteam steht in agilen Projekten sehr im Vordergrund. Darum ist es auch wichtig, zu bewerten, wie das Team zusammenarbeitet, welche Arbeitsmethoden effektiv sind, welche Fähigkeits- und Aufgabenbereiche abgedeckt sind und wie sie sich entwickeln aber auch die individuellen Fähigkeiten der Teammitglieder.

Das Messen von Softskills ist etwas schwieriger, als konkrete Leistungsfakten abzugleichen. Hier gibt es zwei Möglichkeiten:

- 1. Abfragen in Fortschrittsbesprechungen oder bei Team-Retros.
- 2. Indirektes Messen: Beispielsweise Verwendung und Nutzungszeiten von Hilfsmitteln, zeitliche Veränderungen der Produktivität, Anzahl von Teaminteraktionen, etc..



## **BREAKOUT-SESSION**

Gruppenarbeit zum Thema "Machine Learning Cycle" danach weiter im Plenum



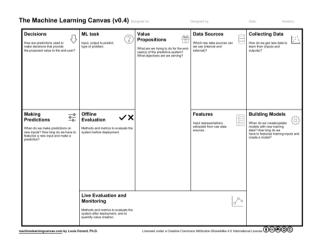
- 1. Wie lässt sich der Fortschritt oder Erfolg Ihres Projektes messen? Definieren Sie einige geeignete Metriken.
- Formulieren Sie zu Ihren Zielen und Aufgaben jeweils ein "Definition of done".

- Die Teilnehmer werden nach ihren Projektgruppen auf Räume verteilt.
- Legen Sie innerhalb der Arbeitsgruppen fest, wer dokumentiert und wer im Anschluss die Ergebnisse vorträgt.





# Der Machine Learning Canvas macht ein gemeinsames Projektverständnis möglich.



https://www.ownml.co/machine-learning-canvas

# **Machine Learning Canvas**

- Strukturiert Projektaufgabe
- Hilft wesentliche Fragestellungen nicht zu übersehen
- Macht Lieferumfänge transparent
- Klärt Erwartungshaltungen in neuem Umfeld
- Werkzeug, um Zusammenarbeit zwischen Business und IT zu erleichtern
- Hintergrund: "You don't make your company successful by buying a bucket of Al. It's not about waving an Al wand -- there's no intrinsic value by itself. It's about picking the right thing to do with Al." - Carl Hillier

Der Machine Learning Canvas integriert 10 Perspektiven.

## Mehrwert

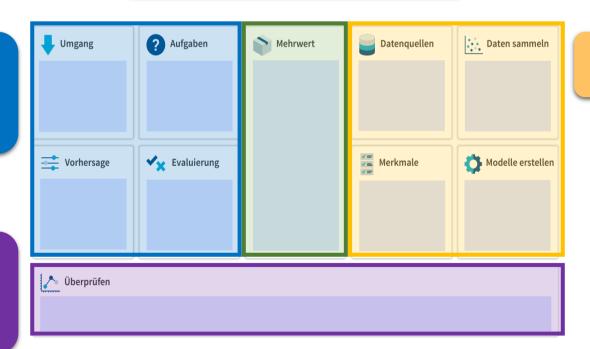
User Stories: Wer - Was - Warum

## Vorhersagen

Wie werden aus Daten Vorhersagen und wie werden sie angewendet?

# Überprüfen

Wie gut ist das Model? Welcher Mehrwert wurde tatsächlich geschaffen?



Lernen



# Beschreibung des Mehrwerts in Form von User-Storys

## "Als [Rolle, wer] möchte ich [Funktion, was] um [Nutzen, warum]."

#### [Rolle] Wer fordert etwas an?

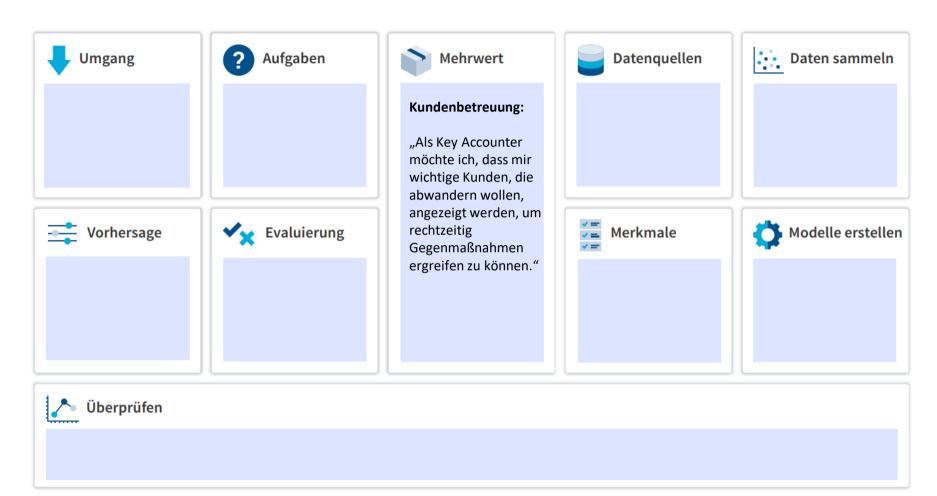
Der Anforderer ist meist der spätere Nutzer des Systems oder der Nutznießer der zu entwickelnden Lösung.

#### [Funktion] Was wünscht sich der Anforderer?

Der Anforderer wünscht sich, dass etwas Bestimmtes passiert. Je klarer und präziser der Wunsch in die User Story formuliert ist, desto hilfreicher ist er für die Realisierung.

#### [Nutzen] Warum ist das wichtig?

Durch die Anforderung soll ein Ziel erreicht werden. Der Anforderer verspricht sich davon einen Nutzen. Die Beschreibung des Grundes der Anforderung liefert weitere Informationen zur richtigen Ausführung.





# Datenquellen – Relevante Daten und Quellentypen

#### Relevante Daten identifizieren

Anfrage von Datenbanken können bei großen Datenmengen viel Zeit in Anspruch nehmen. Es macht Sinn, sich auf die relevanten Daten zu beschränken.

#### Datenquellentypen identifizieren

Je nach Typ der Datenquelle (z.B. interne oder externe Datenbank, APIs, etc.) können Beschränkungen bezüglich des Zugangs, Datenvolumens, u.ä. bestehen, die für den Datenabgriff berücksichtigt werden müssen.

Wenn Daten aus vorgelagerten Verarbeitungsprozessen stammen (z.B. andere ML) müssen diese eventuell erst gestartet werden.





# Sammeln und Aufbereiten von Daten

#### Daten aufbereiten

Häufig können Daten nicht in der Rohfassung für das ML-Modell verwendet werden und müssen erst aufbereitet werden (z.B. beim Überwachten Lernen). Je nach Aufwand sind Kosten und Zeit zu berücksichtigen.

#### Langfristig planen

Wenn ein Vorhersagemodell längere Zeit arbeiten soll, muss es kontinuierlich mit neuen Daten versorgt werden. Frühzeitiges Planen vermeidet Engpässe.

#### **Beeinflusste Daten**

Die Auswahl und das Aufbereiten von Daten beeinflusst das Ergebnis der Vorhersage und kann dieses lenken. Dies sollte stets beachtet werden. Es gibt keine absolut neutrale Methode.









#### **Kundenbetreuung:**

"Als Key Accounter möchte ich, dass mir wichtige Kunden, die abwandern wollen, angezeigt werden, um rechtzeitig Gegenmaßnahmen ergreifen zu können."



## Datenquellen

- CRM
- Zahlungsdatenbank
- Web-Analyse
- Kundenservice
- Emails an den Kunden



# Daten sammeln

- Abonnements: welcher Kunde zahlt nicht mehr
- Schlechte Bewertung des Kundenservice







Merkmale



Modelle erstellen



Überprüfen



# Merkmale – Repräsentative Werte und Transformation

#### Input

Vorhersagen werden anhand von Werten für Objekte ermittelt. Diese müssen entweder numerisch, oder kategorisch ausgedrückt werden. Gibt es viele Features, ist es sinnvoll diese in verwandte Gruppen einzuteilen.

#### **Transformation**

Häufig werden Aggregationen der Rohdaten benötigt. Die Wahl der Aggregation hat einen großen Einfluss auf das Ergebnis der Vorhersage (z.B. Zeitabschnitte).

#### Temporäre Daten

Viele Daten ändern sich im Laufe der Zeit. Es ist zu entscheiden, wie diese Daten zu erfassen sind (Schnappschuss, Zeitverlauf, Mittelwerte, etc.)



Vorhersage







#### Mehrwert

#### **Kundenbetreuung:**

"Als Key Accounter möchte ich, dass mir wichtige Kunden, die abwandern wollen, angezeigt werden, um rechtzeitig Gegenmaßnahmen ergreifen zu können."



## Datenquellen

- CRM
- Zahlungsdatenbank
- Web-Analyse
- Kundenservice
- Fmails an den Kunden



# Daten sammeln

- Abonnements: welcher Kunde zahlt nicht mehr
- Schlechte Bewertung des Kundenservice



## Merkmale

- Kundendaten: Alter. Wohnort, etc.
- Nutzungsinformationen: Logins, Nutzungsdauer
- Kontakt zum Kundenservice
- Benutzte Geräte



#### Modelle erstellen



Überprüfen



# Modelle erstellen und aktualisieren

#### Anzahl der Modelle

Je nachdem, wie spezifisch ein Modell angelegt ist, muss entscheiden werden, wie viele Modelle oder Modellvariationen benötigt werden.

#### Aktualisierungsfrequenz

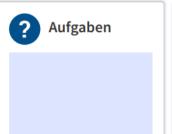
Rahmenbedingungen und Merkmalverhältnisse können sich im Laufe der Zeit ändern. Es sollte festgelegt werden, in welcher Frequenz oder bei welchen Ereignissen die Modelle überprüft und aktualisiert werden.

#### Beschränkungen

Dieser Bereich kann genutzt werden, Beschränkungen für die Modell oder den Generierungsprozess zu dokumentieren (z.B. maximale Größe für Geräte mit wenig Prozessorleistung, maximale Durchlaufzeit, etc.)



- Vorhersage







## Mehrwert

#### **Kundenbetreuung:**

"Als Key Accounter möchte ich, dass mir wichtige Kunden, die abwandern wollen, angezeigt werden, um rechtzeitig Gegenmaßnahmen ergreifen zu können."



## Datenquellen

- CRM
- Zahlungsdatenbank
- Web-Analyse
- Kundenservice
- Fmails an den Kunden



# Daten sammeln

- Abonnements: welcher Kunde zahlt nicht mehr
- Schlechte Bewertung des Kundenservice



## Merkmale

- Kundendaten: Alter. Wohnort, etc.
- Nutzungsinformationen: Logins, Nutzungsdauer
- Kontakt zum Kundenservice
- Benutzte Geräte



### Modelle erstellen

- Einmal pro Monat mit Daten des vorhergehenden Monats
- Die Aktualisierung und Offline-Evaluierung erfolgt über Nacht



Überprüfen



## **BREAKOUT-SESSION**

Gruppenarbeit zum Thema "Machine Learning Canvas" danach weiter im Plenum



Erstellen Sie einen Machine Learning Canvas zu Ihrem Projekt

Teil 1:

Beschreiben Sie die Bereiche Mehrwert und Lernen

- Die Teilnehmer werden nach ihren Projektgruppen auf Räume verteilt.
- Legen Sie innerhalb der Arbeitsgruppen fest, wer dokumentiert und wer im Anschluss die Ergebnisse vorträgt.

# ? Aufgaben

# Aufgabe des Modells – Input und Output

#### Beschreibung der Aufgaben des Modells

Anhand der User Storys, Datenquellen und Merkmalen wird hier kurz zusammenfassend die einzelnen Aufgaben des Modells formuliert.

#### **Input und Output**

Klassifizierung und Erläuterung der Input- und Output-Werte werden hier dargestellt und welche Informationen aus ihnen gewonnen werden können.

Der Modell-Typ wird hier dokumentiert.

#### Vergleichswerte

In manchen Fällen gibt es bereits Vorhersagen oder Schätzungen (Mittelwerte, Expertenschätzungen). Diese können hier beschrieben werden.



#### **Umgang**



## Aufgaben

- Vorhersage auf die Frage "Wird der Kunde im kommenden Monat abwandern?"
- Input: Kunde
- Output: "abwandern" (positive Klasse) oder "nicht abwandern"
- Das ML-Modell ist eine binäre Klassifikation





#### Evaluierung



## Mehrwert

#### **Kundenbetreuung:**

"Als Key Accounter möchte ich, dass mir wichtige Kunden, die abwandern wollen, angezeigt werden, um rechtzeitig Gegenmaßnahmen ergreifen zu können."



## Datenquellen

- CRM
- Zahlungsdatenbank
- Web-Analyse
- Kundenservice
- Fmails an den Kunden



# Daten sammeln

- Ahonnements: welcher Kunde zahlt nicht mehr
- Schlechte Bewertung des Kundenservice



## Merkmale

- Kundendaten: Alter. Wohnort, etc.
- Nutzungsinformationen: Logins, Nutzungsdauer
- Kontakt zum Kundenservice
- Benutzte Geräte



#### Modelle erstellen

- Einmal pro Monat mit Daten des vorhergehenden Monats
- Die Aktualisierung und Offline-Evaluierung erfolgt über Nacht



#### Überprüfen



# Umsetzung der Vorhersagen

#### Vorhersagen in Entscheidungen verwandeln

Welche Aktionen folgen auf die Vorhersage? Wie würde man handeln, wenn man eine 100% Vorhersage bekommen könnte?

#### **Zeitliche Planung**

Wann sollen die Entscheidungen umgesetzt werden. Regelmäßige Termine? Bestehen Abhängigkeiten (Verfügbarkeit der Daten)?

#### Entscheidungen automatisieren

Wie vertrauensvoll sind die Vorhersagen? Kann die Entscheidung automatisch umgesetzt werden oder ist eine Kontrolle notwendig?

#### Vorsicht vor Feedbackschleifen!



- Jeden Monatsanfang anwenden
- Zufällig 25% der Kunden ausfiltern (Vermeidung der Feedbackschleife)
- Verbleibende Kunden nach Abwanderwahrscheinlichkeit und monatlichem Umsatz sortieren
- Kontakt mit den Kunden aufnehmen, die im Aktionsbereich liegen (Schwellenwert)



## Aufgaben

- Vorhersage auf die Frage "Wird der Kunde im kommenden Monat abwandern?"
- Input: Kunde
- Output: "abwandern" (positive Klasse) oder "nicht abwandern"
- Das ML-Modell ist eine binäre Klassifikation



#### Mehrwert

#### **Kundenbetreuung:**

"Als Key Accounter möchte ich, dass mir wichtige Kunden, die abwandern wollen, angezeigt werden, um rechtzeitig Gegenmaßnahmen ergreifen zu können."



## Datenquellen

- CRM
- Zahlungsdatenbank
- Web-Analyse
- Kundenservice
- Fmails an den Kunden



# Daten sammeln

- Ahonnements: welcher Kunde zahlt nicht mehr
- Schlechte Bewertung des Kundenservice



#### - Vorhersage



#### **Evaluierung**



- Kundendaten: Alter. Wohnort, etc.
- Nutzungsinformationen: Logins, Nutzungsdauer
- Kontakt zum Kundenservice
- Benutzte Geräte



#### Modelle erstellen

- Einmal pro Monat mit Daten des vorhergehenden Monats
- Die Aktualisierung und Offline-Evaluierung erfolgt über Nacht





# Vorhersagen planen und durchführen

#### Frequenz und Volumen

Wie viele Modelle müssen in einem definierten Zeitraum durchgeführt werden? Wie ist die zeitliche Verteilung?

#### Zeitlicher Aufwand

Wie lange benötigt der ganze Prozess von Daten sammeln bis Modell rechnen? Wie kann dies im vertretbaren Maße für den Benutzer geplant und umgesetzt werden?

#### Beschränkungen und Optimierungen

Welche Beschränkungen liegen noch für die Ausführung der Vorhersage vor (Rechenleistung, Datenzugang, etc.)? Wie kann die Ausführung den Bedingungen angepasst werden (APIs, reservierte Server, etc.)



- · Jeden Monatsanfang anwenden
- Zufällig 25% der Kunden ausfiltern (Vermeidung der Feedbackschleife)
- · Verbleibende Kunden nach Abwanderwahrscheinlichkeit und monatlichem Umsatz sortieren
- Kontakt mit den Kunden aufnehmen, die im Aktionsbereich liegen (Schwellenwert)



## Aufgaben

- Vorhersage auf die Frage "Wird der Kunde im kommenden Monat abwandern?"
- Input: Kunde
- Output: "abwandern" (positive Klasse) oder "nicht abwandern"
- Das ML-Modell ist eine binäre Klassifikation



#### Mehrwert

#### **Kundenbetreuung:**

"Als Key Accounter möchte ich, dass mir wichtige Kunden, die abwandern wollen, angezeigt werden, um rechtzeitig Gegenmaßnahmen ergreifen zu können."



#### Datenquellen

- CRM
- Zahlungsdatenbank
- Web-Analyse
- Kundenservice
- Fmails an den Kunden



# Daten sammeln

- Abonnements: welcher Kunde zahlt nicht mehr
- Schlechte Bewertung des Kundenservice



#### - Vorhersage

- leden Monat wird die Kundendatenbank neu bearbeitet und eine Vorhersage berechnet
- Die Vorhersage soll über Nacht stattfinden, um tagsüber die IT-Infrastruktur nicht zu stören.



#### **Evaluierung**



# Merkmale

- Kundendaten: Alter. Wohnort, etc.
- Nutzungsinformationen: Logins, Nutzungsdauer
- Kontakt zum Kundenservice
- Benutzte Geräte



#### Modelle erstellen

- Einmal pro Monat mit Daten des vorhergehenden Monats
- Die Aktualisierung und Offline-Evaluierung erfolgt über Nacht





# Prediction Positive Negative True Positive Positive Positive Negative Positive Negative

# Offline die Zuverlässigkeit des Modells überprüfen

#### Zuverlässigkeit prüfen

Wie zuverlässig sind die berechneten Vorhersagen? Ist das Modell stabil?

#### Repräsentative Testdaten verwenden

Die Testdaten sollten in der Zusammensetzung den Daten für Vorhersagen entsprechen. Die Größe sollte so gewählt sein, dass zuverlässig eine Vorhersage berechnet werden kann.

#### KPIs festlegen und Schwellenwerte festlegen

Mit welchen Messwerten lässt sich die Zuverlässigkeit des Modells überprüfen. Was sind sinnvolle Schwellenwerte? Optimierung aber immer nur mit einer Testkategorie durchführen, da sich Kategorien gegenseitig beeinflussen können.



- · Jeden Monatsanfang anwenden
- Zufällig 25% der Kunden ausfiltern (Vermeidung der Feedbackschleife)
- Verbleibende Kunden nach Abwanderwahrscheinlichkeit und monatlichem Umsatz sortieren
- Kontakt mit den Kunden aufnehmen, die im Aktionsbereich liegen (Schwellenwert)



# Aufgaben

- Vorhersage auf die Frage "Wird der Kunde im kommenden Monat abwandern?"
- Input: Kunde
- Output: "abwandern" (positive Klasse) oder "nicht abwandern"
- Das ML-Modell ist eine binäre Klassifikation



#### Mehrwert

#### **Kundenbetreuung:**

"Als Key Accounter möchte ich, dass mir wichtige Kunden, die abwandern wollen. angezeigt werden, um rechtzeitig Gegenmaßnahmen ergreifen zu können."



## Datenquellen

- CRM
- Zahlungsdatenbank
- Web-Analyse
- Kundenservice
- Fmails an den Kunden



# Daten sammeln

- Abonnements: welcher Kunde zahlt nicht mehr
- Schlechte Bewertung des Kundenservice



#### - Vorhersage

- leden Monat wird die Kundendatenbank neu bearbeitet und eine Vorhersage berechnet
- Die Vorhersage soll über Nacht stattfinden, um tagsüber die IT-Infrastruktur nicht zu stören



#### Evaluierung

- Getestet wird mit den Daten des letzten Monats
- KPI: Abwanderungsrate in ML Gruppe muss kleiner als in Kontrollgruppe sein
- KPI: 1 Monat Rückblick. Fehler max. 5%



# Merkmale

- Kundendaten: Alter. Wohnort, etc.
- Nutzungsinformationen: Logins, Nutzungsdauer
- Kontakt zum Kundenservice
- Benutzte Geräte



#### Modelle erstellen

- Einmal pro Monat mit Daten des vorletzten Monats
- Die Aktualisierung und Offline-Evaluierung erfolgt über Nacht





# Modell im Einsatz überwachen

#### Umsetzung im Einsatz prüfen

Lässt sich die Vorhersage wie geplant erstellen? Sind die Daten verfügbar?

#### Zuverlässigkeit anhand von KPIs regelmäßig prüfen

Messwerte und Zeitplan zur Überprüfung definieren. Schwellenwerte festlegen, bei denen das Modell erfolgreich bzw. als nicht mehr zuverlässig gilt.

#### Wenn die Vorhersagen zu sehr abweichen

Lassen sich die Abweichungen ursächlich erklären? Liegt die Abweichung im Datenprozess oder an externen Faktoren? Lässt sich der Datenprozess optimieren bzw. lassen sich die externen Faktoren durch Änderung des Modells zuverlässig vorhersagen?

#### Maßnahmen bei Abweichung definieren!



- · Jeden Monatsanfang anwenden
- Zufällig 25% der Kunden ausfiltern (Vermeidung der Feedbackschleife)
- Verbleibende Kunden nach Abwanderwahrscheinlichkeit und monatlichem Umsatz sortieren
- Kontakt mit den Kunden aufnehmen, die im Aktionsbereich liegen (Schwellenwert)



#### Aufgaben

- Vorhersage auf die Frage "Wird der Kunde im kommenden Monat abwandern?"
- Input: Kunde
- Output: "abwandern" (positive Klasse) oder "nicht abwandern"
- Das ML-Modell ist eine binäre Klassifikation



#### Mehrwert

#### **Kundenbetreuung:**

"Als Key Accounter möchte ich, dass mir wichtige Kunden, die abwandern wollen. angezeigt werden, um rechtzeitig Gegenmaßnahmen ergreifen zu können."



#### Datenquellen

- CRM
- Zahlungsdatenbank
- Web-Analyse
- Kundenservice
- Fmails an den Kunden



#### Daten sammeln

- Ahonnements: welcher Kunde zahlt nicht mehr
- Schlechte Bewertung des Kundenservice



#### - Vorhersage

- leden Monat wird die Kundendatenbank neu bearbeitet und eine Vorhersage berechnet
- Die Vorhersage soll über Nacht stattfinden, um tagsüber die IT-Infrastruktur nicht zu stören



#### **Evaluierung**

- Getestet wird mit den Daten des letzten Monats
- KPI: Abwanderungsrate in ML Gruppe muss kleiner als in Kontrollgruppe sein
- KPI: 1 Monat Rückblick. Fehler max. 5%



# Merkmale

- Kundendaten: Alter. Wohnort, etc.
- Nutzungsinformationen: Logins, Nutzungsdauer
- Kontakt zum Kundenservice
- Benutzte Geräte



#### Modelle erstellen

- Einmal pro Monat mit Daten des vorletzten Monats
- Die Aktualisierung und Offline-Evaluierung erfolgt über Nacht



- KPI: Vergleich der Vorhersage des letzten Monats mit der Vergleichsgruppe (Abwanderungsrate und verlorener Umsatz)
- KPI: Abwanderungsgrate bei den vorhergesagten Kunden
- KPI: ROI mit Differenz Umsatz Modell und Kontrollgruppe abzüglich der Kosten für Modell und Maßnahmen.



# **BREAKOUT-SESSION**

Gruppenarbeit zum Thema "Machine Learning Canvas" danach weiter im Plenum



Erstellen Sie einen Machine Learning Canvas zu Ihrem Projekt

Teil 2:

Beschreiben Sie die Bereiche Vorhersagen und Überprüfen







# **BREAKOUT-SESSION**

Gruppenarbeit zum Thema "Der erste Sprint " danach weiter im Plenum



- Formulieren Sie aus den Projektzielen und dem Machine Learning Canvas eine Liste der Aufgaben für den ersten Sprint.
- Schätzen Sie den Aufwand (in Stunden) für jede Aufgabe und dokumentieren sie ihn.



