



Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences



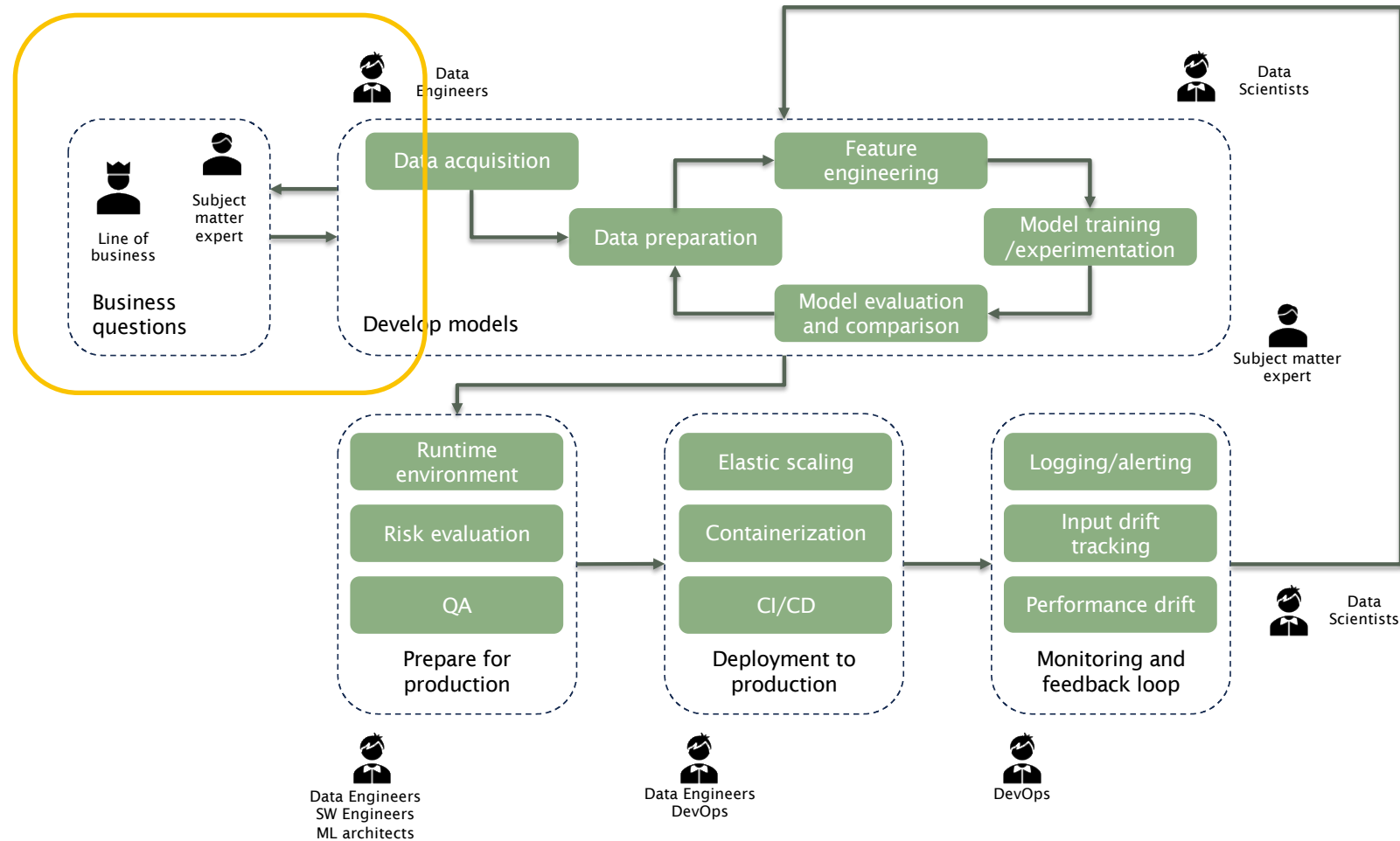
Lebenszyklus von Maschine Lernen Modelle

Untertitel

► Violeta Vogel, TI BFH

-
- ▶ Teil zwei: KI-Anwendungsfälle finden, Geschäftsfragen klären, Daten verstehen

Das realistische Bild eines ML-Lebenszyklus innerhalb einer durchschnittlichen Organisation



-
- ▶ Anwendungsfall für KI finden

Anwendungsfall für KI finden

Ziele der Projekte

- ▶ Das Ziel eines KI-Projekts darf nicht auf einer abstrakten Ebene bleiben. Es sollte einen konkreten Anwendungsfall oder ein geschäftliches Problem adressieren und messbare Ergebnisse sowie einen messbaren ROI aufweisen.
- ▶ Beispiel:
 - ▶ Schlechtes Ziel: Steigerung des Gesamtumsatzes
 - ▶ Gutes Ziel:
 - ▶ Kaufempfehlungen auf Basis von Produkten, die voraussichtlich zusammen gekauft werden
 - ▶ und die durchschnittliche Bestellmenge pro Kunde um X% zu erhöhen.
 - ▶ was sich positiv auf übergeordnete Ziele wie die Steigerung des durchschnittlichen Umsatzes pro Kunde und des Gesamtumsatzes auswirken wird.

Anwendungsfall für KI finden

Kategorien von KI-Anwendungsfällen (nach Komplexität und Wert geordnet)

1. Intelligente Prognosen und Datenanalysen zur Unterstützung verschiedener Entscheidungen
2. Innovative Prozess- oder Serviceautomatisierung zur Kostensenkung und Produktivitätssteigerung
3. Neue Produkte und Dienstleistungen, die einen zusätzlichen Wert generieren
4. Einfachere oder bessere Benutzererfahrung und autonome Systeme (Bots, Roboter, Autos usw.)

Anwendungsfall für KI finden

Beispiele

- ▶ Anwendungsfall bei LATAM Airlines: Verbesserung der Genauigkeit der Treibstoffberechnung, um das Mitführen von zusätzlichem Treibstoff zu vermeiden
- ▶ Bis zu 35 % des Umsatzes von Amazon stammen aus KI-gestützten Empfehlungen.
 - ▶ Durch die Einführung der Empfehlungen *für häufig zusammen gekaufte Artikel* konnte Amazon den durchschnittlichen Warenkorbwert und Bestellwert pro Kunde steigern, was wiederum den durchschnittlichen Umsatz pro Kunde und den E-Commerce-Umsatz von Amazon pro Quartal erhöhte.
- ▶ Netflix schätzt den Wert seiner personalisierten Empfehlungsmaschine auf 1 Milliarde US-Dollar pro Jahr.

Anwendungsfall für KI finden

Die 12 häufigsten Anwendungsfälle für KI

1. Produktempfehlungen
 - ▶ Empfehlungssysteme, die Nutzern Produkte basierend auf ihrem Verhalten, ihrer Kaufhistorie, ihrer Profilsegmentierung und anderen Faktoren anbieten.
2. Chatbots
 - ▶ Chatbots, die mit Nutzern interagieren und ihnen während des gesamten Nutzerprozesses Unterstützung, Anleitung und Hilfestellung bieten.
3. Marketing und Inhalte
 - ▶ Gen AI kann Marketingstrategien und -pläne entwickeln, Wettbewerbsanalysen durchführen und Marketingmaterialien wie Blogbeiträge oder E-Mails erstellen.
4. Kundenstimmungsanalyse
 - ▶ Messung der Gefühle und Meinungen, die Kunden online auf Websites, Foren, in sozialen Medien und anderen Kanälen äußern.

Anwendungsfall für KI finden

Die 12 häufigsten Anwendungsfälle für KI

5. Absatzprognose
 - ▶ Berechnung der Wahrscheinlichkeit von Kundenkäufen, Umsätzen und Konversionen.
6. Preisoptimierung
 - ▶ Durch die Berechnung von Faktoren wie bisherigen Preisen, der Qualität Ihrer Marke, den Preisen der Konkurrenz, den Betriebskosten, der Marktsituation und vielem mehr wird der optimale Preis ermittelt.
7. Cybersicherheit
 - ▶ Angriffe erkennen und vorhersagen
8. Betrugsvorhersage
 - ▶ Transaktionen analysieren, um Bedrohungen in Echtzeit zu erkennen und sie zu blockieren, bevor sie auftreten

Anwendungsfall für KI finden

Die 12 häufigsten Anwendungsfälle für KI

9. Ressourcenoptimierung

- ▶ Wege finden, um Rechenressourcen effizienter zu nutzen, um Kosten zu senken und Nachhaltigkeit zu fördern

10. Bedarfsprognose

- ▶ Eine präzise Bedarfsprognose und die Überwachung der Produktion helfen, Verschwendung zu vermeiden.

11. Gesundheitspflege

- ▶ Vorhersage von Krankheitszuständen und Verschlechterung des Patientenzustands sowie Unterstützung bei Behandlung, Medikamentengabe und Triage auf Intensivstationen.

12. Vorausschauende Wartung

- ▶ Störungen frühzeitig erkennen, um Zeit zu sparen und den Betrieb aufrechtzuerhalten.

Übung 15 Minuten

ZWILLING

KPT

Inselspital

1. Überlegen Sie sich alle möglichen Anwendungsfälle für dieses Unternehmen (erste PowerPoint-Folie).

-
- ▶ So starten Sie Ihr KI-Projekt

Geschäftsfragen

Hauptaufgaben

1. Geschäftsziele festlegen
2. Situation einschätzen
3. Ziele im Bereich Data Science / Data Engineering festlegen
4. Projektplan erstellen

Geschäftsfragen

Aufgabe 1: Geschäftsziele ermitteln

- ▶ Das erste Ziel des Datenanalysten ist es, aus geschäftlicher Sicht genau zu verstehen, was der Kunde wirklich erreichen will.
- ▶ Häufig hat der Kunde viele konkurrierende Ziele und Einschränkungen, die angemessen gegeneinander abgewogen werden müssen.
- ▶ Ziel des Analysten ist es, von Anfang an wichtige Faktoren aufzudecken, die den Ausgang des Projekts beeinflussen können.
- ▶ Eine mögliche Folge der Vernachlässigung dieses Schrittes ist, dass man viel Aufwand betreibt, um die richtigen Antworten auf die falschen Fragen zu finden.

Geschäftsfragen

Ergebnisse von Aufgabe 1: Geschäftsziele ermitteln

► **Hintergrund:**

- Erfassen Sie die Informationen, die zu Beginn des Projekts über die Geschäftslage der Organisation bekannt sind.

► **Geschäftsziele:**

- Beschreiben Sie das Hauptziel des Kunden aus geschäftlicher Sicht. Zusätzlich zum primären Geschäftsziel gibt es typischerweise weitere damit verbundene Fragen, die der Kunde beantworten möchte. Beispielsweise könnte das primäre Geschäftsziel darin bestehen, bestehende Kunden zu halten, indem vorhergesagt wird, wann sie zur Konkurrenz wechseln könnten. Beispiele für damit verbundene Fragen sind: „Wie beeinflusst der primär genutzte Kanal (z. B. Geldautomat, Filialbesuch, Internet) die Kundenbindung?“ oder „Würden niedrigere Geldautomatengebühren die Abwanderung wertvoller Kunden deutlich reduzieren?“

Geschäftsfragen

Ergebnisse von Aufgabe 1: Geschäftsziele ermitteln

► **Kriterien für den Geschäftserfolg:**

- Beschreiben Sie die Kriterien für ein erfolgreiches oder nützliches Projektergebnis aus betriebswirtschaftlicher Sicht. Diese können sehr spezifisch und objektiv messbar sein, beispielsweise die Reduzierung der Kundenabwanderung auf ein bestimmtes Niveau, oder sie können allgemein und subjektiv sein, wie etwa „nützliche Einblicke in die Beziehungen gewinnen“. Im letzteren Fall sollte angegeben werden, wer die subjektive Beurteilung vornimmt.

Übung 15 Minuten

ZWILLING

KPT

Inselspital

1. Überlegen Sie sich alle möglichen Anwendungsfälle für dieses Unternehmen.
(Erste PowerPoint-Folie)
2. Wählen Sie einen Anwendungsfall aus und bestimmen Sie die Geschäftsziele.
(zweite PowerPoint-Folie)

Geschäftsfragen

Aufgabe 2: Situation einschätzen

- ▶ Diese Aufgabe beinhaltet eine detailliertere Recherche zu allen Ressourcen, Einschränkungen, Annahmen und anderen Faktoren, die bei der Festlegung des Ziels der Datenanalyse und des Projektplans berücksichtigt werden sollten.
- ▶ In der vorherigen Aufgabe ging es darum, schnell zum Kern der Situation vorzudringen. Hier sollen Sie die Details ausführlicher darstellen.

Geschäftsfragen

Ergebnisse von Aufgabe 2: Situation einschätzen

▶ **Ressourceninventar:**

- ▶ Liste der für das Projekt verfügbaren Ressourcen:
 - ▶ einschließlich Personal (Business-Experten, Datenexperten, technischer Support, Data-Science-Experten),
 - ▶ Daten (festgelegte Auszüge, Zugriff auf Live-, gespeicherte oder operative Daten),
 - ▶ Rechenressourcen (Hardwareplattformen),
 - ▶ und Software (Data-Science-Tools) sonstige relevante Software).

Geschäftsfragen

Ergebnisse von Aufgabe 2: Situation einschätzen

► Anforderungen, Annahmen und Einschränkungen

- Führen Sie alle Projektanforderungen auf, einschließlich Fertigstellungstermin, Verständlichkeit und Qualität der Ergebnisse sowie Sicherheits- und Rechtsaspekte. Stellen Sie im Rahmen dieser Dokumentation sicher, dass Sie die Daten verwenden dürfen.
- Führen Sie die Annahmen des Projekts auf. Dies können Annahmen über die Daten sein, die sich im Rahmen des Data-Science überprüfen lassen, aber auch nicht überprüfbare Annahmen über das mit dem Projekt verbundene Geschäft. Letztere sollten besonders dann aufgeführt werden, wenn sie die Gültigkeit der Ergebnisse beeinflussen.
- Listen Sie die Einschränkungen des Projekts auf. Dies können Einschränkungen hinsichtlich der Verfügbarkeit von Ressourcen sein, aber auch technologische Einschränkungen wie die Größe des Datensatzes, der für die Modellierung praktisch verwendet werden kann.

Geschäftsfragen

Ergebnisse von Aufgabe 2: Situation einschätzen

► **Risiken und Eventualitäten**

- Listen Sie die Risiken oder Ereignisse auf, die das Projekt verzögern oder zum Scheitern bringen könnten.
- Führen Sie die entsprechenden Notfallpläne auf und beschreiben Sie, welche Maßnahmen ergriffen werden, falls diese Risiken oder Ereignisse eintreten.

Geschäftsfragen

Ergebnisse von Aufgabe 2: Situation einschätzen

► Terminologie

- Erstellen Sie ein Glossar der für das Projekt relevanten Terminologie. Dieses kann zwei Komponenten umfassen:
 - Ein Glossar relevanter Geschäftsbegriffe, das zum Geschäftsverständnis des Projekts beiträgt. Die Erstellung dieses Glossars ist eine sinnvolle Maßnahme zur Wissensgewinnung und Weiterbildung.
 - Ein Glossar der Data-Science-Terminologie, illustriert mit Beispielen, die für das jeweilige Geschäftsproblem relevant sind.

Übung 15 Minuten

ZWILLING

KPT

Inselspital

1. Überlegen Sie sich alle möglichen Anwendungsfälle für dieses Unternehmen. (Erste PowerPoint-Folie)
2. Wählen Sie einen Anwendungsfall aus und bestimmen Sie die Geschäftsziele. (zweite PowerPoint-Folie)
3. Anforderungen, Annahmen und Einschränkungen auflisten. (dritte PowerPoint-Folie)

Geschäftsfragen

Ergebnisse von Aufgabe 2: Situation einschätzen

► **Kosten und Nutzen**

- Erstellen Sie eine Kosten-Nutzen-Analyse für das Projekt, in der Sie die Projektkosten den potenziellen Nutzen für das Unternehmen im Erfolgsfall gegenüberstellen. Der Vergleich sollte so detailliert wie möglich sein. Verwenden Sie beispielsweise in einem kommerziellen Kontext monetäre Kennzahlen.
- ROI berechnen: Kapitalrendite

Geschäftsfragen

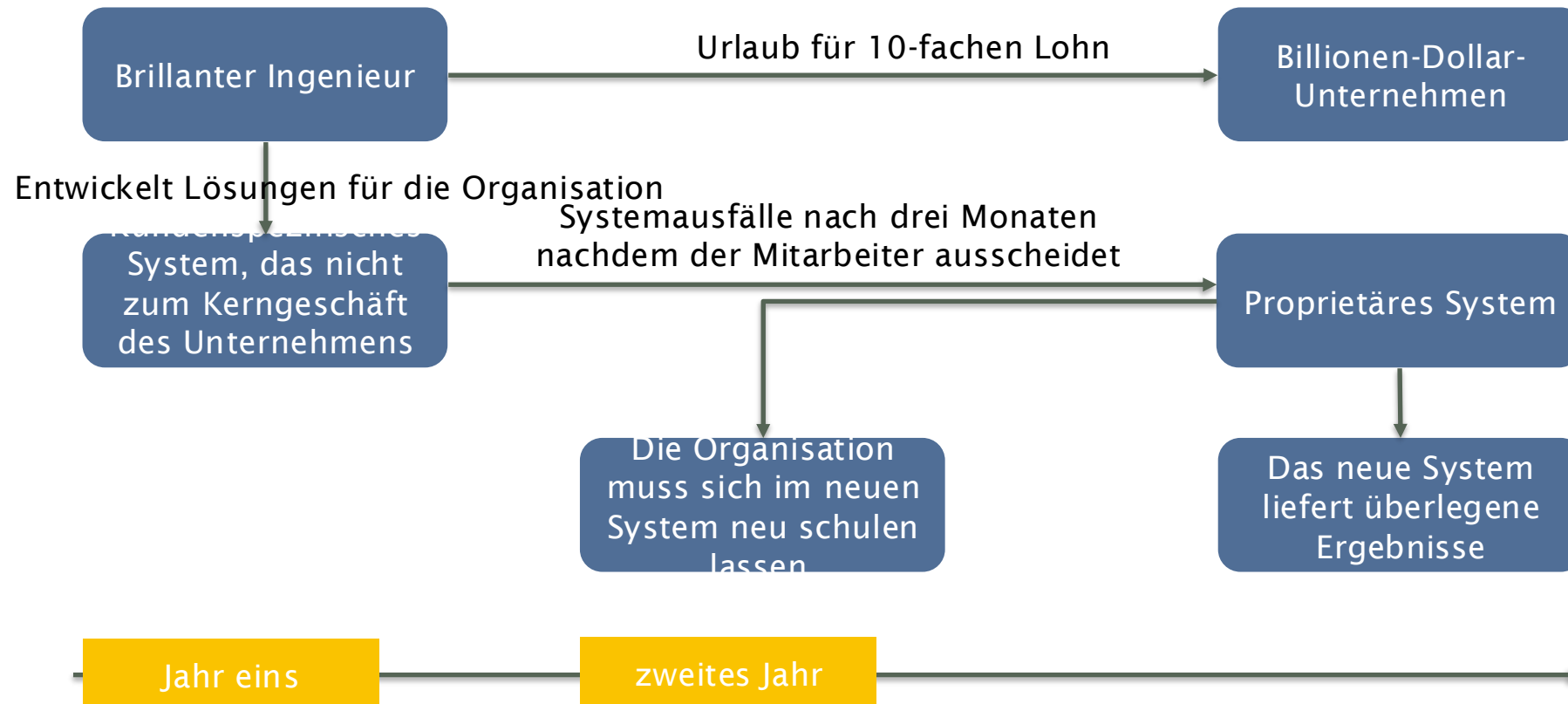
Den ROI (Return on Investment) von Unternehmenslösungen verstehen

Die beste Wahl ist nicht die „kostenlose“, sondern die Lösung, die den höchsten ROI liefert, da dieser ROI die Geschwindigkeit des profitablen Unternehmens erhöht.

1: Kein Wert, keine Kosten, kein ROI	2: Hoher Wert, hohe Kosten, negativer ROI	3: Hoher Wert, niedrige Kosten, positiver ROI
Wert = 0	Wert = 1 Mio.	Wert = 1 Mio.
Kosten = 0	Kosten = 2 Mio.	Kosten = 0,5 Mio.
ROI = 0	ROI = -1 Mio.	ROI = 0,5 Mio.

Geschäftsfragen

ROI (Return on Investment) bei Unternehmenslösungen verstehen: Systemdilemma



Geschäftsfragen

Berechnung des ROI des Projekts

▶ **Investitionen:**

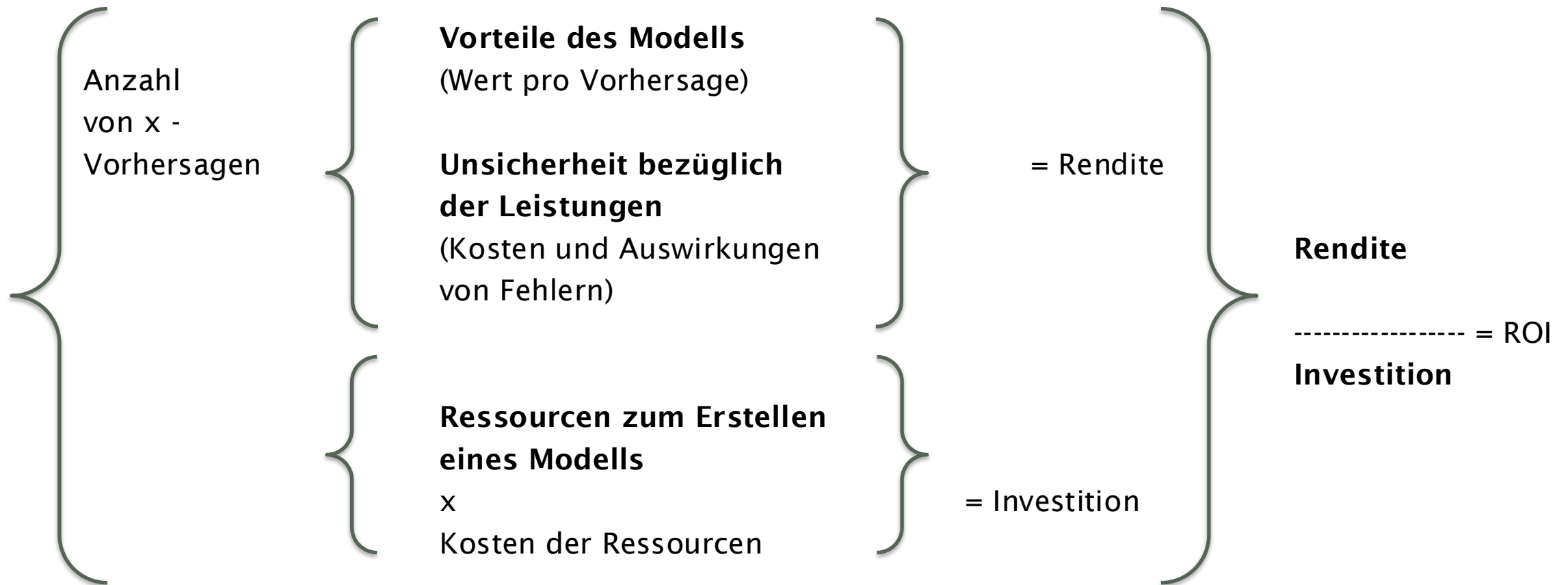
- ▶ Personen (Data Scientists, Data Engineers, MLOps usw.)
- ▶ Rechen- und Dateninfrastruktur
- ▶ Softwarelizenzen und Dienstleistungen
- ▶ Berater und Schulungen

▶ **Rückgabe:**

- ▶ Kosteneinsparungen
- ▶ Erhöhte Einnahmen
- ▶ Zeitersparnis oder gesteigerte Produktivität
- ▶ Erhöhte Wettbewerbsfähigkeit oder Nutzerbasis

Geschäftsfragen

Berechnung des ROI des Projekts



Geschäftsfragen

Berechnung des ROI des ML-Modells: Gewinn pro Vorhersage

- ▶ Der Gewinn pro Vorhersage wird anhand folgender Formel geschätzt:

$$\hat{a} = (a - (1 - l) * e)$$

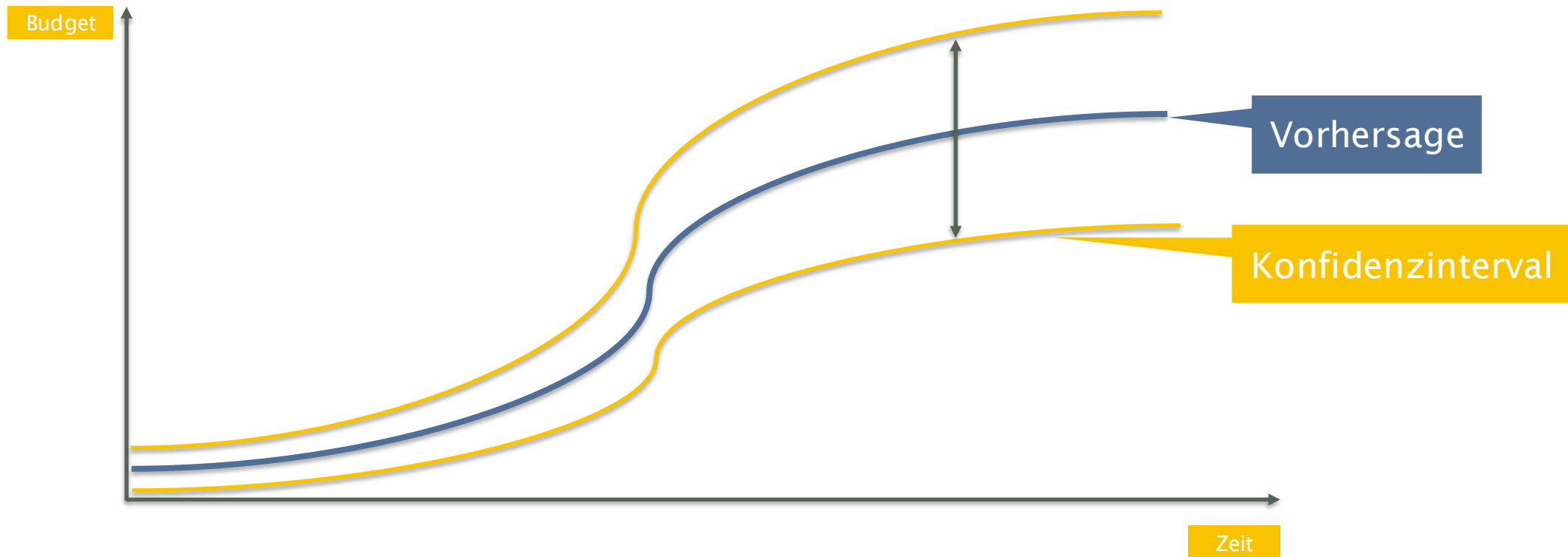
- ▶ \hat{a} - bereinigte Einsparung (Gewinn pro Prognose)
- ▶ a - erwartete Einsparung (Wert pro Prognose)
- ▶ l - berechnete durchschnittliche Genauigkeit (diese erhalten wir durch das Training eines Modells)
- ▶ e - Kosten der manuellen Fehlerkorrektur (Unsicherheit des Nutzens)

- ▶ Um die bereinigten Einsparungen zu ermitteln, müssen wir das Verhältnis der falsch prognostizierten Werte ($1 - \text{Genauigkeit}$) zu den Kosten eines Fehlers berücksichtigen. Die bereinigten Einsparungen geben die tatsächlichen Einsparungen nach Abzug der Fehleranzahl an.

Geschäftsfragen

Berechnung des ROI des ML-Modells

- Was wäre, wenn der Algorithmus Empfehlungen zur Zuverlässigkeit seiner Vorhersage abgeben könnte, um das Risiko zu reduzieren?



Geschäftsfragen

Berechnung des ROI des ML-Modells

- ▶ Die Schritte zur Entwicklung eines robusteren Schätzalgorithmus sind nur geringfügig komplexer als die der vorherigen Gleichung. Dies sind die Schritte:
 - ▶ Auswahl des Schwellenwerts χ , der die Vorhersagen mit hoher und niedriger Konfidenz trennt
 - ▶ Führe den Algorithmus aus und erhalte Vorhersagen.
 - ▶ Berechnen Sie den Konfidenzwert für jede Vorhersage.
 - ▶ Die unsichersten und kostspieligsten Einträge werden so herausgefiltert, dass der Aufteilungsschwellenwert χ^2 erfüllt ist.
 - ▶ Die Vorhersagen werden anhand des benutzerdefinierten Aufteilungsschwellenwerts aufgeteilt.
 - ▶ Berechnen Sie aus dem verbleibenden Datensatz den Konfidenzgenauigkeitswert: \hat{I}
 - ▶ Wenden Sie die untenstehende Gleichung an, um die angepassten Ersparnisse zu erhalten: \hat{a}

Geschäftsfragen

Berechnung des ROI des ML-Modells

$$\hat{a} = ((a - (1 - \hat{I}) * e) * \kappa) - ((1 - \kappa) * \hat{e})$$

- ▶ Wobei \hat{a} die bereinigte Einsparung (Gewinn pro Prognose) bezeichnet.
- ▶ a stellt die erwartete Einsparung dar.
- ▶ \hat{I} ist die berechnete durchschnittliche Konfidenzgenauigkeit,
- ▶ e sind die Kosten für die manuelle Korrektur eines Fehlers.
- ▶ κ ist der Schwellenwert für die Inhaftierung von Nutzern,
- ▶ und \hat{e} sind die Kosten der manuellen Überprüfung.
- ▶ Die obige Gleichung gibt die angepassten Einsparungen \hat{a} nach dem Entfernen von Vorhersagen mit geringer Konfidenz und dem Hinzufügen der Kosten für die manuelle Überprüfung zurück, die vom Split-Threshold beeinflusst werden: $(1 - \kappa)$.

Geschäftsfragen

Berechnung des ROI des ML-Modells

- ▶ Wie legt man den Aufteilungsschwellenwert χ fest?
 - ▶ Das hängt von der Genauigkeit des Algorithmus ab.
 - ▶ Falls Ihr Algorithmus eine hohe Genauigkeit aufweist (in diesem Fall gehen wir von einer Genauigkeit von 100 Prozent aus), dann ist der rot markierte Teil der obigen Gleichung nicht erforderlich und kann weggelassen werden.
 - ▶ Hinweis: Der grün markierte Teil liefert dasselbe Ergebnis wie die erste Gleichung.
 - ▶ In der Realität ist die Genauigkeit meist deutlich geringer, weshalb wir die Vorhersagen in zwei Segmente aufteilen müssen. Falls der Algorithmus eine geringere Genauigkeit liefert, müssen wir den Anteil der manuellen Überprüfung erhöhen und das Vertrauen in den grün markierten Bereich deutlich reduzieren. In diesem Beispiel verwenden wir ein Aufteilungsverhältnis von 90/10, d. h. 90 Prozent der Vorhersagen werden als verlässlich angesehen und 10 Prozent manuell überprüft.

Geschäftsfragen: Berechnung des ROI eines ML-Modells: Beispiel

Wiederaufnahme ins Krankenhaus

- ▶ Die Kosten für Wiedereinweisungen ins Krankenhaus machen einen großen Teil der Ausgaben für stationäre Krankenhausleistungen aus. Diabetes zählt nicht nur zu den zehn häufigsten Todesursachen weltweit, sondern ist auch die teuerste chronische Erkrankung in den Vereinigten Staaten.
- ▶ Diabetespatienten, die stationär behandelt werden, haben ein höheres Risiko für eine Wiedereinweisung als andere Patienten. Daher birgt die Senkung der Wiedereinweisungsraten bei Diabetespatienten ein großes Potenzial zur erheblichen Reduzierung der Gesundheitskosten.
- ▶ Der verwendete Datensatz stammt vom Center for Machine Learning and Intelligent Systems der University of California, Irvine, und umfasst über 100.000 Attribute und 50 Merkmale. Der Datensatz ist auf Kaggle verfügbar.

Wiederaufnahme ins Krankenhaus

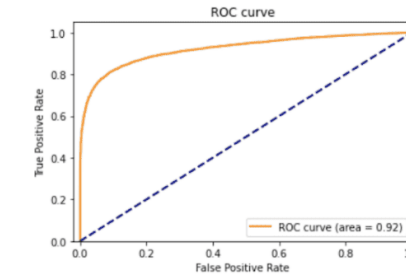
- ▶ Es kann jede beliebige Klassifikationsbibliothek verwendet werden.
- ▶ das XGBoost -Klassifikationsmodell verwendet.
- ▶ Die Trainingsdaten umfassen 91 589

$$\begin{aligned} a &= 5 \text{ minutes saved} & \hat{a} &= (a - (1 - I) * e) \\ e &= 35 \text{ minutes} & \hat{a} &= (5 - (1 - 0.86) * 35) \\ I &= 0.86 & \hat{a} &= 0.099 \end{aligned}$$

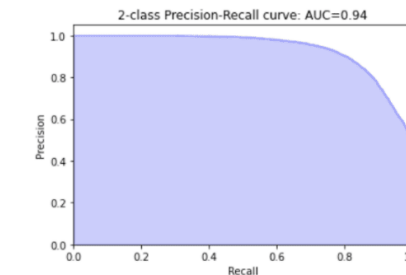
Dadurch können wir mit einer Zeitersparnis von 0,099 Minuten pro Antrag rechnen. Durch die Bearbeitung von 10177 Anträgen können wir 17 Arbeitsstunden einsparen.

Accuracy: 0.8586

AUC Score (Train): 0.921817



Average precision-recall score: 0.94



Precision score: 0.880742

Recall score: 0.825171

Kappa score: 0.716899

Matthews score: 0.718296

Neg Log Loss score: 0.350779

Classification Report					
	precision	recall	f1-score	support	
0	0.84	0.89	0.86	9159	
1	0.88	0.83	0.85	8923	
accuracy			0.86	18082	
macro avg	0.86	0.86	0.86	18082	
weighted avg	0.86	0.86	0.86	18082	

Confusion Matrix
[[8162 997]
[1560 7363]]

Übung 15 Minuten

ZWILLING

KPT

Inselspital

1. Überlegen Sie sich alle möglichen Anwendungsfälle für dieses Unternehmen. (Erste PowerPoint-Folie)
2. Wählen Sie einen Anwendungsfall aus und bestimmen Sie die Geschäftsziele. (zweite PowerPoint-Folie)
3. Anforderungen, Annahmen und Einschränkungen auflisten. (dritte PowerPoint-Folie)
4. Listen Sie die Investitionen und Erträge Ihres ML-Projekts auf (vierte PowerPoint-Folie).

Geschäftsfragen

Aufgabe 3: Ziele für Data Science / Data Engineering festlegen

- ▶ Ein *Unternehmensziel* formuliert die angestrebten Ziele in der Sprache der Geschäftswelt.
- ▶ Ein *Data-Science*-Ziel formuliert die Projektziele in technischen Begriffen.
 - ▶ Ein Geschäftsziel könnte beispielsweise lauten: „Steigerung der Katalogverkäufe an Bestandskunden“. Ein Ziel im Bereich Data Science könnte lauten: „Vorhersage, wie viele Artikel ein Kunde kaufen wird, basierend auf seinen Käufen der letzten drei Jahre, demografischen Informationen (Alter, Gehalt, Wohnort usw.) und dem Preis des Artikels.“

Geschäftsfragen

Ergebnisse von Aufgabe 3: Bestimmung der Ziele für Data Science / Data Engineering

▶ **Ziele der Datenanalyse**

- ▶ Beschreiben Sie die angestrebten Ergebnisse des Projekts, die zur Erreichung der Geschäftsziele beitragen.

▶ **Erfolgskriterien für Data Science**

- ▶ Definieren Sie die Kriterien für einen erfolgreichen Projektabschluss in technischen Begriffen – zum Beispiel ein bestimmtes Maß an Vorhersagegenauigkeit oder ein Kaufneigungsprofil mit einem bestimmten Grad an „Lift“.
- ▶ Ähnlich wie bei den Kriterien für den Geschäftserfolg kann es erforderlich sein, diese in subjektiven Begriffen zu beschreiben. In diesem Fall sollte die Person oder die Personen, die das subjektive Urteil fällen, benannt werden.

Übung 15 Minuten

ZWILLING

KPT

Inselspital

1. Überlegen Sie sich alle möglichen Anwendungsfälle für dieses Unternehmen. (Erste PowerPoint-Folie)
2. Wählen Sie einen Anwendungsfall aus und bestimmen Sie die Geschäftsziele. (zweite PowerPoint-Folie)
3. Anforderungen, Annahmen und Einschränkungen auflisten. (dritte PowerPoint-Folie)
4. Listen Sie die Investitionen und Erträge Ihres ML-Projekts auf (vierte PowerPoint-Folie).
5. Listen Sie Ihre Ziele für die Datenanalyse auf (fünfte Folie).

Geschäftsfragen

Aufgabe 4: Projektplan erstellen

1. Beschreiben Sie den geplanten Ansatz zur Erreichung der Data-Science-Ziele und damit der Geschäftsziele.
2. Der Plan sollte die im weiteren Verlauf des Projekts durchzuführenden Schritte, einschließlich der anfänglichen Auswahl von Werkzeugen und Techniken, genau beschreiben.

Geschäftsfragen

Ergebnisse von Aufgabe 4: Projektplan erstellen

► **Projektplan:**

- Listen Sie die im Projekt auszuführenden Phasen auf, zusammen mit ihrer Dauer, den benötigten Ressourcen, den Inputs, Outputs und Abhängigkeiten.
 - Wo immer möglich, sollten die umfangreichen Iterationen im Data-Science-Prozess explizit dargestellt werden – beispielsweise Wiederholungen der Modellierungs- und Evaluierungsphasen.
- Im Rahmen der Projektplanung ist es außerdem wichtig, die Abhängigkeiten zwischen Zeitplan und Risiken zu analysieren.
 - Die Ergebnisse dieser Analysen sollten im Projektplan explizit festgehalten werden, idealerweise mit Maßnahmen und Empfehlungen, falls sich die Risiken manifestieren.

Geschäftsfragen

Ergebnisse von Aufgabe 4: Projektplan erstellen

► **Projektplan:**

- Der Projektplan ist ein dynamisches Dokument, da am Ende jeder Phase eine Überprüfung des Fortschritts und der erreichten Ergebnisse erforderlich ist und eine entsprechende Aktualisierung des Projektplans empfohlen wird. Spezifische Überprüfungspunkte für diese Aktualisierungen sind Bestandteil des Projektplans.

Geschäftsfragen

Ergebnisse von Aufgabe 4: Projektplan erstellen

1. Zielsetzung: Was sind die Ziele dieses KI-Fallbeispiels?
2. KPIs: Was qualifiziert für Erfolg und wie wird er gemessen?
3. Daten: Sind genügend Daten vorhanden, um das Modell zu trainieren?
4. Personen: Wer wird für das Projekt verantwortlich sein?
5. Algorithmen: Welchen KI-Ansatz planen Sie zu verwenden?
6. Ethik und Risiken: Gibt es in diesem Anwendungsfall ethische oder rechtliche Probleme?
7. Infrastruktur: Welche technologischen und infrastrukturellen Herausforderungen und Anforderungen gibt es?
8. Kontinuität: Können Sie die Anwendung kontinuierlich überwachen und warten? Können Sie Modell und Daten häufig genug aktualisieren? Wie überprüfen Sie die KPIs und den ROI nach der Bereitstellung der Anwendung?

Übung 15 Minuten

ZWILLING

KPT

Inselspital

1. Überlegen Sie sich alle möglichen Anwendungsfälle für dieses Unternehmen. (Erste PowerPoint-Folie)
2. Wählen Sie einen Anwendungsfall aus und bestimmen Sie die Geschäftsziele. (zweite PowerPoint-Folie)
3. Anforderungen, Annahmen und Einschränkungen auflisten. (dritte PowerPoint-Folie)
4. Listen Sie die Investitionen und Erträge Ihres ML-Projekts auf (vierte PowerPoint-Folie).
5. Listen Sie Ihre Ziele für die Datenanalyse auf (fünfte Folie).
6. Erstellen Sie einen übergeordneten Projektplan (sechste Folie).

Übung 30 Minuten

ZWILLING

KPT

Inselspital

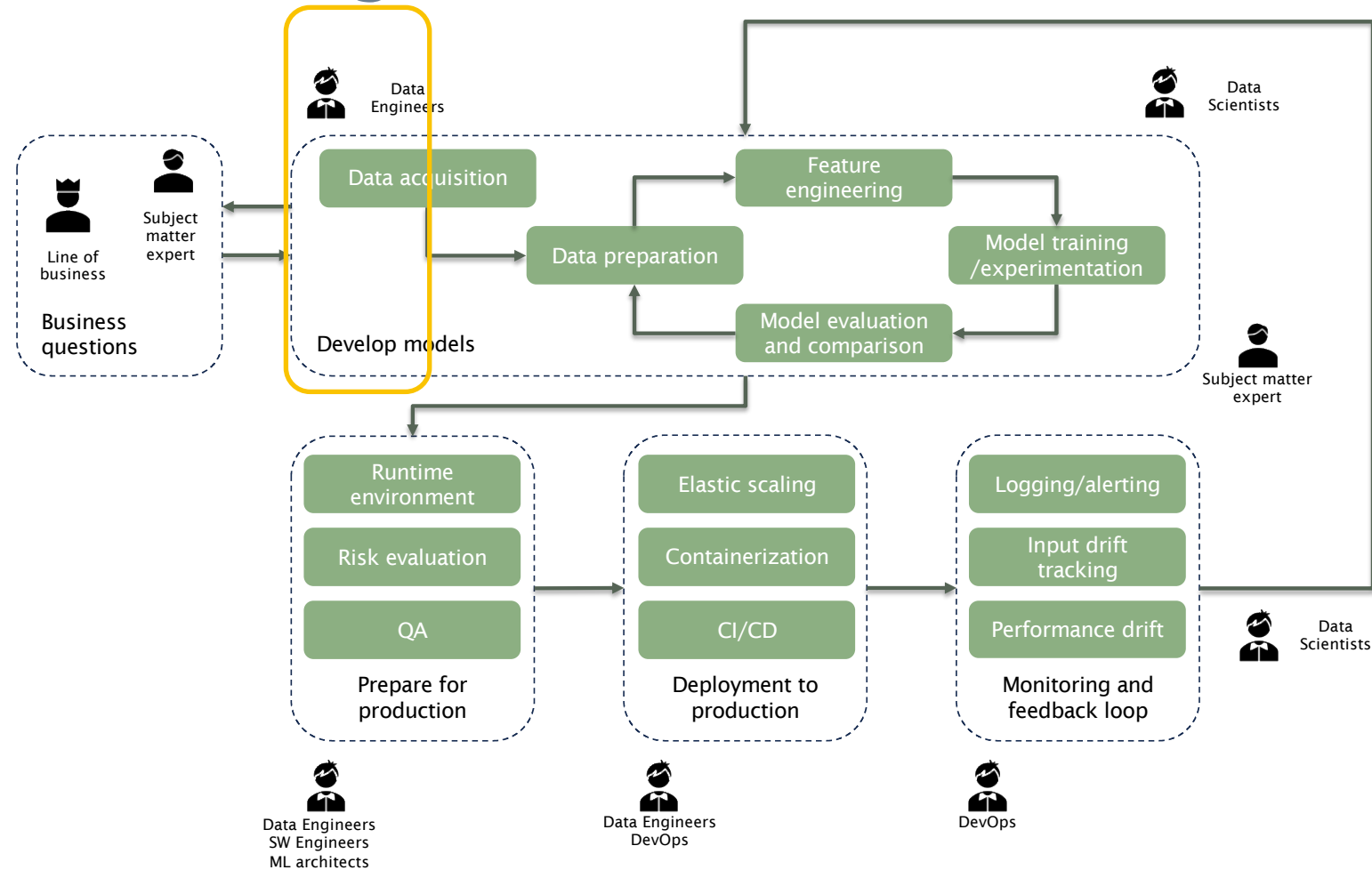
1. Präsentation der Ergebnisse

Geschäftsfragen

Was Sie wissen müssen

1. Nennen Sie mindestens 5 Anwendungsfälle für KI.
2. Welche Investitionen sind für ein ML-Projekt erforderlich? Welche Renditen sind zu erwarten?
3. Wie berechnet man den ROI eines ML-Projekts?
4. Welche Komponenten umfasst ein ML-Projekt?

Das realistische Bild eines ML-Lebenszyklus innerhalb einer durchschnittlichen Organisation



Datenverständnis

Hauptaufgaben

1. Sammeln Sie erste Daten
2. Daten beschreiben
3. Daten erkunden
4. Datenqualität prüfen

Datenverständnis

Aufgabe 1: Sammeln von Anfangsdaten

- ▶ Beschaffen Sie sich die in den Projektressourcen aufgeführten Daten (oder den Zugang zu den Daten).
- ▶ Diese erste Datenerfassung umfasst das Laden von Daten, sofern dies für das Datenverständnis erforderlich ist.
 - ▶ Wenn Sie beispielsweise ein bestimmtes Tool zur Datenanalyse verwenden, ist es sinnvoll, Ihre Daten in dieses Tool zu laden. Dies kann erste Schritte zur Datenaufbereitung nach sich ziehen.
 - ▶ Hinweis: Wenn Sie mehrere Datenquellen verwenden, stellt die Integration ein zusätzliches Problem dar, entweder hier oder in der späteren Datenaufbereitungsphase.

Datenverständnis

Ergebnis von Aufgabe 1: Sammeln von Anfangsdaten

- ▶ **Erster Datenerfassungsbericht (Datenkatalog)**
 - ▶ Listen Sie die erfassten Datensätze zusammen mit ihren Speicherorten, den Methoden, mit denen sie erfasst wurden, und allen aufgetretenen Problemen auf.
 - ▶ Dokumentieren Sie aufgetretene Probleme und deren Lösungen. Dies erleichtert die zukünftige Wiederholung dieses Projekts oder die Durchführung ähnlicher Projekte.

Datenverständnis

Aufgabe 2: Daten beschreiben

- ▶ **Daten beschreiben**

- ▶ Untersuchen Sie die „groben“ oder „oberflächlichen“ Eigenschaften der erfassten Daten und berichten Sie über die Ergebnisse.

Datenverständnis

Ergebnis von Aufgabe 2: Daten beschreiben

- ▶ **Datenbeschreibungsbericht (Datenkatalog)**
 - ▶ Beschreiben Sie die erhobenen Daten, einschließlich:
 - ▶ das Format der Daten,
 - ▶ die Datenmenge (zum Beispiel die Anzahl der Datensätze und Felder in jeder Tabelle),
 - ▶ die Identitäten der Felder,
 - ▶ und alle anderen Oberflächenmerkmale, die entdeckt wurden.
 - ▶ Prüfen Sie, ob die erhobenen Daten die relevanten Anforderungen erfüllen.

Datenverständnis

Aufgabe 3: Daten erkunden

▶ **Daten erkunden**

- ▶ Diese Aufgabe befasst sich mit Fragestellungen des Data Science mithilfe von Abfrage-, Visualisierungs- und Berichtstechniken. Dazu gehören:
 - ▶ Verteilung der Schlüsselattribute (zum Beispiel des Zielattributs einer Vorhersageaufgabe)
 - ▶ Beziehungen zwischen Paaren oder kleinen Anzahlen von Attributen,
 - ▶ Ergebnisse einfacher Aggregationen,
 - ▶ Eigenschaften signifikanter Teilpopulationen,
 - ▶ und einfache statistische Analysen

Datenverständnis

Ergebnisse von Aufgabe 3: Daten erkunden

► **Datenexplorationsbericht**

- Beschreiben Sie die Ergebnisse dieser Aufgabe, einschließlich erster Erkenntnisse oder anfänglicher Hypothesen und deren Auswirkungen auf den weiteren Projektverlauf.
- Fügen Sie gegebenenfalls Grafiken und Diagramme hinzu, um Datenmerkmale aufzuzeigen, die eine weitere Untersuchung interessanter Datenteilmengen nahelegen.

Datenverständnis

Aufgabe 4: Datenqualität überprüfen

- ▶ Prüfen Sie die Qualität der Daten und beantworten Sie dabei Fragen wie:
 - ▶ Sind die Daten vollständig (decken sie alle erforderlichen Fälle ab)?
 - ▶ Ist es korrekt, oder enthält es Fehler, und falls es Fehler gibt, wie häufig sind diese?
 - ▶ Gibt es fehlende Werte in den Daten?
 - ▶ Falls ja, wie werden sie dargestellt, wo kommen sie vor und wie verbreitet sind sie?

Datenverständnis

Ergebnis von Aufgabe 4: Datenqualität überprüfen

▶ **Datenqualitätsbericht**

- ▶ Listen Sie die Ergebnisse der Datenqualitätsprüfung auf;
- ▶ Falls Qualitätsprobleme bestehen, listen Sie mögliche Lösungen auf.
 - ▶ Die Lösung von Datenqualitätsproblemen hängt im Allgemeinen stark von Daten- und Geschäftskenntnissen ab.