

# Präzision und Verlässlichkeit für Data Science Projekte!

## Azenes Data Science Blog

Zug, 10.März 2024 | Version 100 mvi



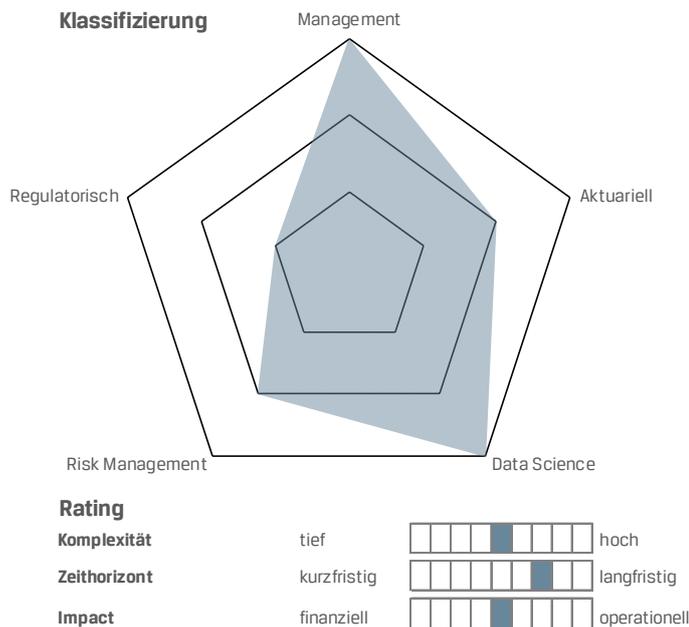
[creativecommons.org](https://creativecommons.org)



## Management Summary

Gleichermassen, wie Data-Science-Projekte in der Versicherungsbranche zunehmend an Bedeutung für entscheidende Geschäftsprozesse gewinnen, entstehen Herausforderungen bei der internen Überprüfung und dem Wissenstransfer zu den Personen mit Entscheidungsbefugnis. Die externe Validierung bietet hier eine effektive Lösung, indem sie Präzision und Verlässlichkeit der Projektergebnisse verbessert und eine fundierte Entscheidungsfindung unterstützt. Sie dient als Brücke zwischen den technischen Teams und dem Management und stellt sicher, dass die Projektergebnisse klar und verständlich kommuniziert werden. Unser Ansatz der externen Validierung schafft durch interdisziplinäre Expertise wertvolle Einsichten für das weitere Projektmanagement und eine solide Basis für interne Audits.

## Azenes Rating für diesen Artikel



Die Artikel von Azenes werden folgendermassen bewertet: Im oberen Teil wird die Relevanz für verschiedene Bereiche aufgezeigt, im unteren Teil wird die Implementation bewertet.

## 1 Problemspezifikation

In der Versicherungswelt stützen sich vermehrt wesentliche operationelle, taktische und strategische Entscheidungen auf der Grundlage von Data-Science-Projekten. Dazu zählen Elemente wie Tarifgestaltung, Risikoeinschätzung, Betrugserkennung, Schadensregulierung und die Bindung von Kund\_innen.

Ein vorherrschendes Problem ist, dass diese Data-Science-Projekte im Vergleich zu anderen Geschäftsfeldern seltener einer umfassenden Überprüfung durch unabhängige Einrichtungen unterzogen werden. Hinzu kommt, dass internen Audit-Abteilungen oft das notwendige Fachwissen fehlt, um solche Überprüfungen wirksam durchzuführen.

Ebenfalls stellen wir fest, dass wichtige Erkenntnisse und Resultate der Data-Science-Projekte bei der Übersetzung von den Fachabteilungen zu den entscheidenden Gremien verloren gehen.

## 2 Lösung

### 2.1 Sicherheit durch externe Validierung

Eine effektive Lösung für die oben identifizierten Probleme ist die Einbindung einer externen Validierung (siehe Abbildung 1) in ein Data-Science-Prozess. Einerseits kann durch eine externe Validierung die Projektgenauigkeit verstärkt werden. Die resultierenden, präziseren und zuverlässigeren Ergebnisse führen zu einer höheren Effizienz und Effektivität in allen Phasen eines Data-Science-Prozesses. Andererseits unterstützt die externe Validierung eine informierte Entscheidungsfindung. Im Idealfall kann eine externe Unterstützung die Brücke zwischen internen Data-Science-Teams und dem Management schlagen. Eine neutrale und kompetente Vermittlerin stellt sicher, dass die Kommunikation klar und effektiv ist. Somit wird eine kohärente und zielgerichtete Strategieumsetzung gewährleistet.

Zusätzlich bieten die Ergebnisse und Berichte unserer externen Validierung eine solide Basis, auf die sich interne Audit-Abteilungen stützen können. Dies erleichtert es, die internen Audits präziser und zielgerichteter durchzuführen, indem sich die Audit-Abteilungen auf verifizierte und detaillierte Informationen aus den Validierungs-Berichten beziehen. Somit wird eine umfassende und effektive Überprüfung der Data-Science-Projekte innerhalb der Organisation ermöglicht.

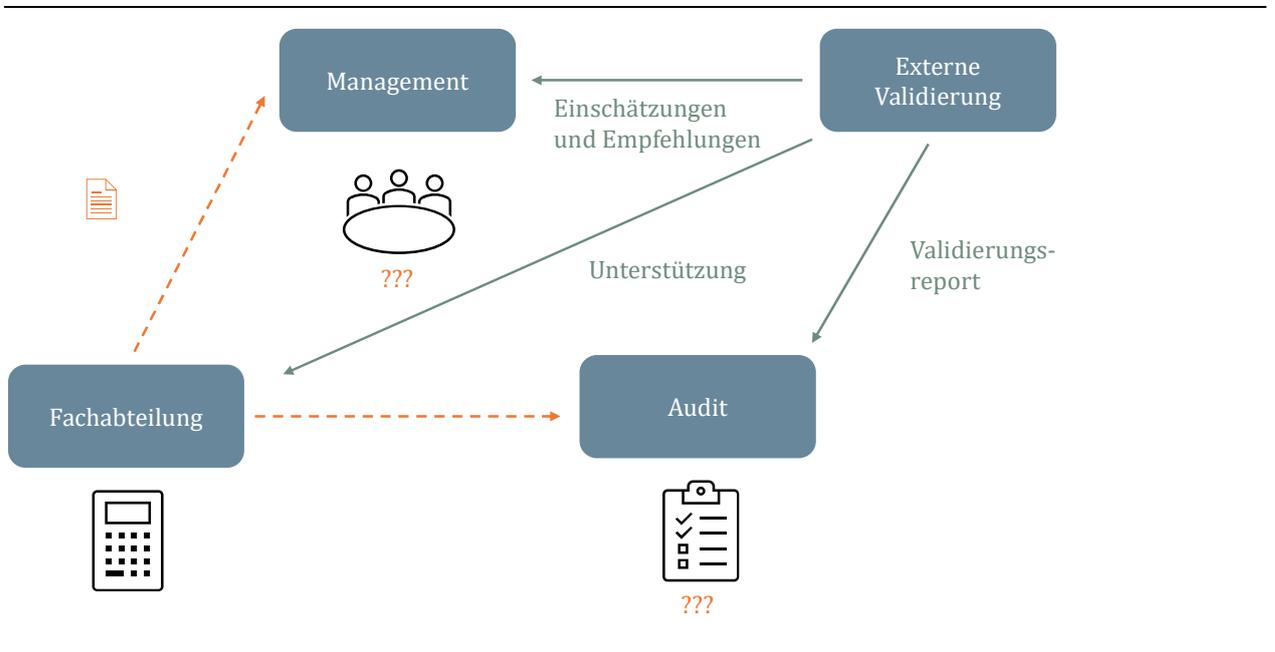


Abbildung 1: Eine externe Validierung schafft Sicherheit und verbindet alle wichtigen Stellen.

Die Validierung eines Data-Science-Projekts ist ein zutiefst interdisziplinäres Unterfangen, das ein breites Spektrum an Fachwissen und Erfahrungen erfordert. Es verlangt nicht nur umfassende Kenntnisse in der Durchführung von Validierungsprojekten, sondern auch ein tiefes Verständnis für strategische Fragestellungen und eine hohe Expertise im Bereich Data Science. Diese Kombination ist entscheidend, da sie eine ganzheitliche Bewertung ermöglicht, die sowohl die technischen Aspekte des Data-Science-Projekts als auch die strategischen und geschäftlichen Auswirkungen berücksichtigt.

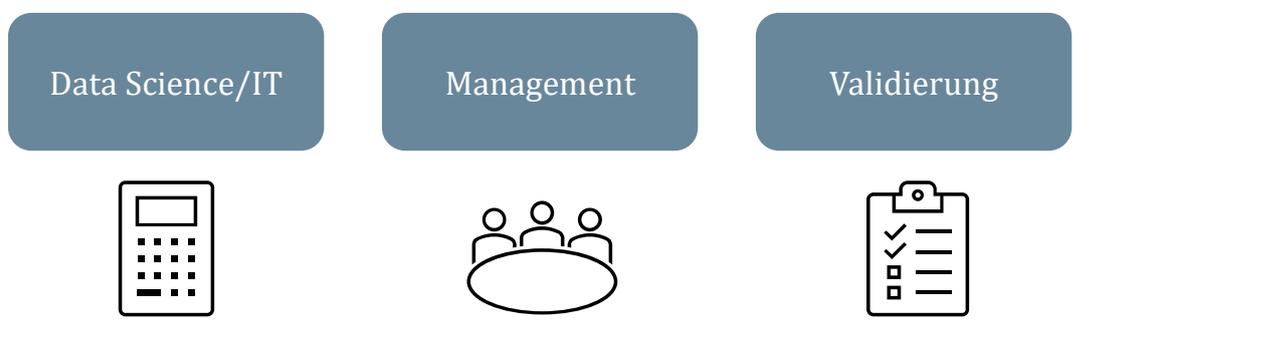


Abbildung 2: Involvierte Disziplinen einer Validierung eines Data-Science Projekts.

Im nachfolgenden Abschnitt präsentieren wir einen Validierungsansatz, mit dem sichergestellt wird, dass oben genannte Mehrwerte auch realisiert werden.

## 2.2 Ein Validierungsansatz

Ein früherer [Blogartikel](#) konnte die Wichtigkeit eines sauberen Prozesses für Data Science Projekte herauschälen (siehe Abbildung 3). Wir orientieren uns bei unserem Validierungsansatz an einem solchen Prozess und bearbeiten die Teilschritte «Strategie», «Umsetzung» und «Finalisierung» als einzelne Validierungsmodule. Diese Module werden im Folgenden einzeln erläutert.

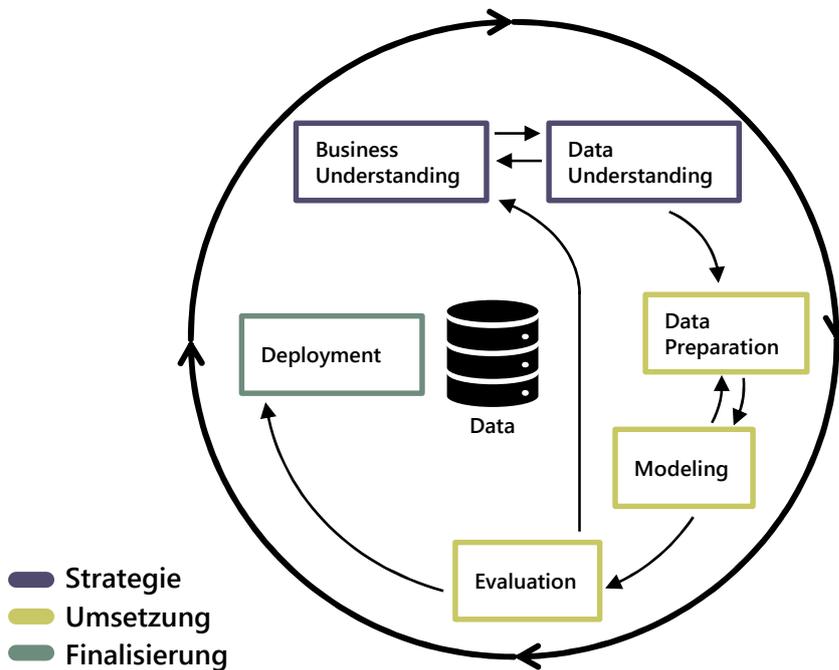


Abbildung 3: Cross-Industry Standard Process für Data Mining (CRISP-DM), eigene Illustration und Erweiterungen

### 2.2.1 Validierungsmodul «Strategie»

Abbildung 4 zeigt die drei grundlegenden Themenbereiche, die im Validierungsmodul «Strategie» behandelt werden.



Abbildung 4: Themen im Validierungsmodul «Strategie»

### **A.1 Ausrichtung und Ziel**

Dieser Bereich konzentriert sich auf die grundlegende Ausrichtung und Zielsetzung des Projekts. Hierbei geht es um:

- **Bedürfnisanalyse:** Verstehen, welche spezifischen Bedürfnisse oder Probleme das Projekt adressieren soll. Dies beinhaltet die Identifizierung der Stakeholder und deren Erwartungen.
- **Projektziel:** Klare Definition der Ziele, die mit dem Projekt erreicht werden sollen. Dies umfasst sowohl kurzfristige als auch langfristige Ziele.
- **Erfolgskriterien:** Festlegung messbarer Kriterien, anhand derer der Erfolg des Projekts bewertet wird.
- **Risikoanalyse:** Identifikation potenzieller Risiken, die das Projekt beeinträchtigen könnten, einschliesslich technischer, operationeller und finanzieller Risiken.

### **A.2 Operationelle Grundlagen**

Dieser Abschnitt befasst sich mit den operativen Aspekten des Projekts:

- **Ist-Zustand:** Analyse des aktuellen Zustands, einschliesslich bestehender Systeme, Prozesse und Datenstrukturen.
- **Datenbasis:** Bewertung der verfügbaren Datenquellen, Datenqualität und der Eignung der Daten für das Projektziel.
- **Ressourcen:** Einschätzung der zur Verfügung stehenden Ressourcen (bspw. Budget, Personal und technische Infrastruktur).
- **Know-how:** Beurteilung der vorhandenen Kompetenzen und Fähigkeiten innerhalb des Teams und gegebenenfalls Identifizierung von Schulungs- oder Einstellungsbedarf.

### **A.3 Planung**

Dieser Teil bezieht sich auf die Strukturierung und das Management des Projekts:

- **Projektplan:** Entwicklung eines detaillierten Plans, der die Schritte zur Zielerreichung aufzeigt, einschliesslich Zeitrahmen und Ressourcenzuweisung.
- **Meilensteine:** Festlegung von Schlüsselpunkten im Projekt, an denen bestimmte Ziele erreicht sein sollten. Diese dienen als Kontrollpunkte zur Überwachung des Fortschritts.
- **Verantwortlichkeiten:** Klare Zuweisung von Aufgaben und Verantwortlichkeiten innerhalb des Teams. Dies gewährleistet, dass jedes Teammitglied seine Rolle und seine Beiträge zum Projekt kennt.

Ein Beispiel, wie ein Questionnaire für das Modul «Strategie» aussehen könnte, ist in Abbildung 5 ersichtlich.

Check	Rating	Comment
A.11 Wurden klare Bedürfnisse formuliert?	1.-5.	[some info]
A.12 Wurden Projektziele definiert?		
A.13 Stimmen Projektziele mit der Geschäftsstrategie überein?		
A.14 Sind Erfolgskriterien erfasst worden?		
A.15 Wie sind die Erfolgskriterien einzuschätzen		
A.16 Wurden mögliche Risiken oder Probleme des Projekts erfasst?		
...		
A.21 Wurde Ist-Zustand erfasst?	1.-5.	[some info]
A.22 Sind Daten für ein erfolgreiches Projekt verfügbar/geeignet?		
A.23 Sind Einschätzungen und Dokumentationen zu folgenden Themen vorhanden? Datensammlung/-beschreibung/-exploration, Qualität, Sicherheit ...		
A.24 Wurden die vorhandenen Ressourcen und das Know-how kontrolliert und dokumentiert?		
...		
A.31 Existiert ein Projektplan?	1.-5.	[some info]
A.32 Werden Ressourcen im Zeitplan alloziert?		
A.33 Ist der Zeitplan realistisch?		
A.34 Gibt es klar definierte Zuweisungen von Aufgaben und Teilschritten.		
...		

Abbildung 5: Validation «Strategie» – Questionnaire Beispiel

### 2.2.2 Validierungsmodul «Umsetzung»

Abbildung 6 zeigt die drei grundlegenden Themenbereiche, die im Validierungsmodul «Umsetzung» behandelt werden.

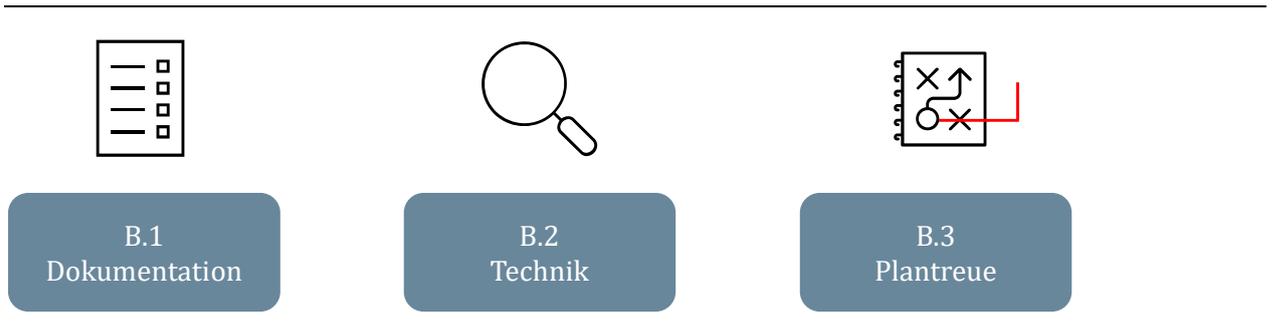


Abbildung 6: Themen im Validierungsmodul «Umsetzung»

#### B.1 Dokumentation

Dieser Bereich konzentriert sich auf die Überprüfung und Aufzeichnung der Projektumsetzung:

- **Methodik:** Detaillierte Aufzeichnung der verwendeten Methoden und Prozesse, um es anderen Teams zu ermöglichen, ähnliche Analysen oder Experimente durchzuführen.

- **Replikation:**
  - Datenaufbereitung: Detaillierte Aufzeichnung aller Schritte der Datenaufbereitung, um die Nachvollziehbarkeit und Transparenz des Prozesses zu gewährleisten.
  - Ergebnisse: Sicherstellung wie die Ergebnisse des Projekts bei wiederholter Anwendung unter gleichen Bedingungen konsistent bleiben.
- **Resultate:** Erfassung und Analyse der erzielten Ergebnisse im Vergleich zu den ursprünglich festgelegten Zielen und Erfolgskriterien. Dies beinhaltet eine Bewertung der Effektivität und Genauigkeit der Ergebnisse.
- **Unzulänglichkeiten:** Identifikation und Dokumentation von Schwachstellen, Herausforderungen oder Problemen, die während der Projektumsetzung aufgetreten sind und deren Auswirkungen auf das Projekt.
- **Entscheidungsgremien:**
  - Kommunikation: Sicherstellung, dass alle relevanten Informationen, Ergebnisse und Empfehlungen effektiv an die Entscheidungsgremien kommuniziert werden. Dies beinhaltet die Präsentation der Ergebnisse in einer verständlichen und handlungsorientierten Form.
  - Feedback: Aktive Einbindung der Entscheidungsgremien, um deren Rückmeldungen und Einschätzungen zum Projektverlauf und den Ergebnissen zu erhalten. Dies fördert das Verständnis und die Akzeptanz der Ergebnisse.

### **B.2 Analysetechniken**

In diesem Abschnitt wird die Angemessenheit und Wirksamkeit der verwendeten Techniken bewertet:

- **Technikevaluation:**
  - Übereinstimmung mit akademischen und praktischen Standards: Überprüfung, ob die verwendeten Techniken und Methoden sowohl den aktuellen akademischen Standards als auch den bewährten Praktiken in der Industrie entsprechen.
  - Effektivität und Eignung: Bewertung der Effektivität und Angemessenheit der eingesetzten Techniken in Bezug auf die Projektziele und -bedingungen.
- **Korrektheit der Umsetzung:**
  - Methodenkonformität: Überprüfung, ob die Implementierung der Techniken gemäss den vorgesehenen Methoden und Spezifikationen erfolgt ist.
  - Fehlerprüfung: Identifikation und Analyse von Fehlern oder Inkonsistenzen in der Umsetzung, die die Ergebnisse beeinträchtigen könnten.
- **Korrektheit der Resultatinterpretation:**
  - Angemessenheit der Schlussfolgerungen: Bewertung, ob die gezogenen Schlussfolgerungen aus den Daten und Ergebnissen angemessen und begründet sind.
  - Vermeidung von Fehlinterpretationen: Identifikation von potenziellen Missverständnissen oder Fehlinterpretationen in der Analyse und Ergebnisdarstellung.

### **B.3 Plantreue**

Dieser Teil beschäftigt sich mit der Einhaltung oder der Anpassung des ursprünglichen Projektplans:

- **Übereinstimmung mit dem Projektplan:** Bewertung, inwieweit die tatsächliche Projektumsetzung dem aktuellen Plan und den festgelegten Meilensteinen entspricht.
- **Abweichungen:** Analyse von Abweichungen vom aktuellen Plan, deren Gründe und Auswirkungen auf das Projektziel.
- **Anpassungen:** Anpassung des Projektplans auf die neuen Gegebenheiten.
- **Verantwortlichkeitsprüfung:** Überprüfung, ob die zugewiesenen Verantwortlichkeiten und Rollen eingehalten wurden und wie effektiv die Teammitglieder ihre Aufgaben erfüllt haben.

Ein Beispiel, wie ein Questionnaire für das Modul «Umsetzung» aussehen könnte, ist in Abbildung 7 ersichtlich.

Check	Rating	Comment
B.11 Gibt es eine Dokumentation zu folgenden Schritten: - Daten säubern/aufbereiten/transformieren/integrieren/formatieren - Modellselektion/-implementierung - Evaluationsmethoden/-prozess/-benchmarking	1.-5.	[some info]
B.12 Ist eine Replikation der Umsetzung möglich?		
B.13 Können die Resultate repliziert werden?		
B.14 Existiert eine Dokumentation zu möglichen Schwächen und Erweiterungen?		
B.15 Werden die Entscheidungsgremien über die Resultate informiert?		
B.16 Werden Inputs der Entscheidungsgremien eingebunden?		
...		
B.21 Erfüllen die gewählten Techniken die Standards, um die Projektziele zu erreichen?	1.-5.	[some info]
B.22 Erfüllen die gewählten Techniken die nötigen akademischen und praktischen Standards?		
B.23 Ist die Umsetzung korrekt (bspw. Codes)?		
B.24 Werden aus den Resultaten die richtigen Schlüsse gezogen?		
B.25 Wurden alternative Techniken besprochen?		
...		
B.31 Wird nach dem erstellten Plan vorgegangen?	1.-5.	[some info]
B.32 Werden allfällige Änderungen der Vorgehensweise (neue Iteration im Data Science Prozess) im Plan festgehalten?		
B.33 Werden die zugewiesenen Rollen eingehalten?		
...		

Abbildung 7: Validation «Umsetzung» – Questionnaire Beispiel

### 2.2.3 Validierungsmodul «Finalisierung»

Abbildung 8 zeigt die drei grundlegenden Themenbereiche, die im Validierungsmodul «Finalisierung» behandelt werden.

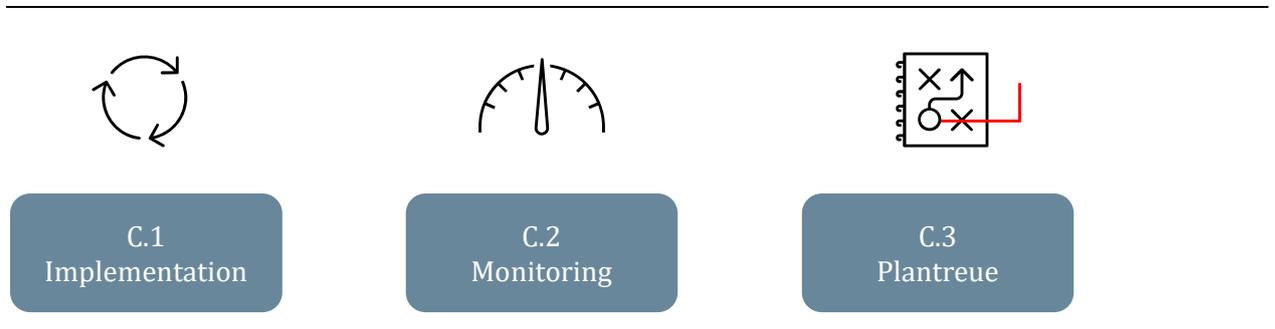


Abbildung 8: Themen im Validierungsmodul «Finalisierung»

### **C.1 Implementation**

Dieser Teil bezieht sich auf die saubere Überführung vom Teilschritt «Umsetzung» in den Teilschritt «Finalisierung»:

- **Überprüfung der Implementierungsprozesse:** Es muss sichergestellt werden, dass die Umsetzung der Data-Science-Lösungen gemäss dem Projektplan erfolgt.
- **Validierung der Systemintegration:** Die nahtlose Integration der entwickelten Modelle und Algorithmen in die bestehenden Systeme muss bestätigt werden.
- **Endbenutzer\_innen-Training:** Bewertung, ob angemessene Schulungsmassnahmen für die Endbenutzer\_innen durchgeführt wurden, um eine effektive Nutzung der neuen Tools und Systeme zu gewährleisten.
- **Dokumentation und Wissensmanagement:** Es sollte überprüft werden, ob alle notwendigen Dokumentationen erstellt und so aufbereitet wurden, dass sie für die relevanten Stakeholder zugänglich und verständlich sind.

### **C.2 Monitoring**

Hierbei geht es um die kontinuierliche Überwachung und Bewertung der Projektergebnisse:

- **Leistungsüberwachung:** Regelmässige Überprüfung der Performance und des Nutzens der implementierten Lösungen. Dies umfasst die Überwachung von KPIs und anderer Erfolgsmetriken.
- **Langzeitbewertung:** Beurteilung der Langzeiteffekte und Nachhaltigkeit der umgesetzten Lösungen.
- **Anpassungsbedarf:** Identifikation von Bereichen, die weiterer Anpassungen oder Verbesserungen bedürfen, basierend auf den Monitoring-Ergebnissen.

### **C.3 Plantruee**

Die Überprüfung der Übereinstimmung des Projekts mit dem ursprünglichen Plan:

- **Plan-Realitäts-Abgleich:** Bewertung, inwieweit die finalisierten Ergebnisse und Prozesse mit dem aktuellen Projektplan übereinstimmen.

- **Umgang mit Abweichungen:** Analyse der Gründe für eventuelle Abweichungen vom Plan und Bewertung, ob diese Abweichungen zu einer Verbesserung oder Verschlechterung des Projektergebnisses geführt haben.
- **Lernerfahrungen:** Ableitung von Lernerfahrungen aus den Abweichungen und deren Integration in zukünftige Projekte oder Prozessverbesserungen.

Ein Beispiel, wie ein Questionnaire für das Modul «Finalisierung» aussehen könnte, ist in Abbildung 7 ersichtlich.

Check	Rating	Comment
C.11 Wurde geprüft, wie sich die entwickelten Modelle in die bestehenden Systemlandschaften und Prozesse integrieren?	1.-5.	[some info]
C.12 Wurde die Implementierung gemäss den getroffenen Entscheidungen durchgeführt?		
C.13 In welchem Masse wurden Endbenutzer und Stakeholder in die Nutzung eingeführt und entsprechend geschult?		
C.14 Ist die Dokumentation der implementierten Lösungen vollständig, aktuell und für alle Stakeholder verständlich?		
...		
C.21 Wie effektiv sind die etablierten Monitoring-Systeme bei der Überwachung der implementierten Lösungen?	1.-5.	[some info]
C.22 Inwieweit ermöglichen die gesammelten Daten eine präzise Analyse und frühzeitige Erkennung von Problemen?		
C.23 Wie gut funktionieren die Feedback-Mechanismen und inwiefern tragen sie zur kontinuierlichen Verbesserung bei?		
C.24 Wie schnell und effektiv können aus dem Monitoring gewonnene Erkenntnisse umgesetzt werden?		
...		
C.31 Wie genau entspricht die finalisierte Umsetzung den im Projektplan festgelegten Zielen und Zeitplänen?	1.-5.	[some info]
C.32 Wie gut werden Abweichungen vom Projektplan erkannt, analysiert und kommuniziert?		
C.33 Wie effektiv sind die entwickelten Strategien im Umgang mit Planabweichungen?		
C.34 In welchem Masse werden die Lernerfahrungen aus Planabweichungen dokumentiert und für zukünftige Projekte genutzt?		
...		

Abbildung 9: Validation «Finalisierung» – Questionnaire Beispiel

#### 2.2.4 Weitere mögliche Module

Neben den grundlegenden Modulen, die sich durch den CRISP-DM aus Abbildung 3 ergeben, können weitere Module oder Modulvertiefungen eine umfassendere und tiefgreifendere Validierung des Data-Science-Projekts ermöglichen. Hier einige Vorschläge für zusätzliche Module und Vertiefungen:

##### Infrastruktur

- Bewertung der vorhandenen IT-Infrastruktur im Hinblick auf die Anforderungen des Data-Science-Projekts.

- Planung und Implementierung notwendiger Upgrades oder Erweiterungen der Systemlandschaft.

#### ***Datensicherheit und Datenschutz***

- Überprüfung der Sicherheitsmassnahmen zum Schutz der Daten vor unbefugtem Zugriff oder Verlust.
- Validierung der Implementierung von Verschlüsselung, Zugriffssteuerung und anderen Sicherheitstechnologien.
- Sicherstellung der Einhaltung von Datenschutzgesetzen und -richtlinien.
- Durchführung von Datenschutz-Folgenabschätzungen und regelmässigen Audits.

#### ***Ethik und Compliance***

- Bewertung der ethischen Implikationen von Data-Science-Anwendungen.
- Überprüfung der Compliance mit internen Richtlinien und externen Vorschriften.

#### ***Endbenutzer\_innenakzeptanz und Change Management***

- Untersuchung der Akzeptanz der Data-Science-Lösungen bei den Endbenutzer\_innen.
- Entwicklung und Umsetzung von Change-Management-Strategien zur Förderung der Akzeptanz.

#### ***Performance und Skalierbarkeit***

- Bewertung der Leistungsfähigkeit der Data-Science-Modelle unter verschiedenen Bedingungen.
- Überprüfung der Skalierbarkeit der Lösungen im Hinblick auf zukünftige Anforderungen und Datenwachstum.

#### ***Nachhaltigkeit und Langzeitwirkung***

- Analyse der Nachhaltigkeit der implementierten Lösungen hinsichtlich Ressourceneinsatz und langfristiger Wartbarkeit.
- Bewertung der langfristigen strategischen Auswirkungen des Projekts auf das Unternehmen.

#### ***Kund\_innen- und Marktanalyse***

- Vertiefung der Einsichten in Kund\_innenverhalten und Marktbedingungen durch erweiterte Data-Science-Modelle.
- Anpassung der Data-Science-Strategie basierend auf Kund\_innenfeedback und Markttrends.

#### ***Innovationsmanagement***

- Bewertung der Fähigkeit des Projekts, Innovationen zu fördern und neue Geschäftsmodelle oder -prozesse zu unterstützen.
- Integration von Forschung und Entwicklung in das Projekt, um fortlaufend innovative Ansätze zu integrieren.

Jedes dieser Validierungs-Module trägt dazu bei, weitere Aspekte eines Data-Science-Projekts zu überprüfen und kann helfen, die Robustheit und Zukunftsorientierung der

Data-Science-Initiativen zu stärken. Durch die Einbeziehung dieser Module in die Validierung kann ein ganzheitlicher und nachhaltiger Ansatz gefördert werden, der über traditionelle Methoden hinausgeht und eine breitere Perspektive auf das Gesamtprojekt bietet.

### 2.2.5 Validierungs-Spider

Der «Azenes Validierungs-Spider» in Abbildung 10 bietet eine visuelle Darstellung der Ergebnisse einer Validierung, die wieder auf den Data-Science-Prozess aus Abbildung 3 angewendet werden. Diese Art der Darstellung ist in ihrer Einfachheit und Klarheit besonders wertvoll, um komplexe Sachverhalte auf den ersten Blick verständlich zu machen.

Die Grafik zeigt deutlich, dass der Bereich «Business Understanding» eine hohe Bewertung erhalten hat, was auf eine starke Leistung in diesem Aspekt der Validierung hindeutet. Andere Bereiche, wie «Data Preparation», liegen im mittleren Bereich, was auf mögliche Gefahren hinweist. Ein Bereich, hervorgehoben durch die rote Linie, zeigt eine tiefe Bewertung, was darauf hindeutet, dass in diesem Bereich erheblicher Handlungsbedarf besteht und das Erreichen der Projektziele mit grosser Wahrscheinlichkeit nicht möglich ist.

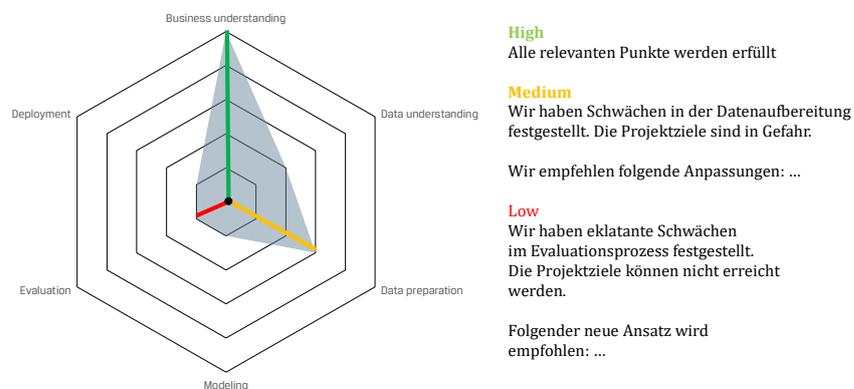


Abbildung 10: Übersichtlicher Auswertungs-Spider der Data-Science-Validierung.

Die klare Visualisierung ist für die Validierung eines Data-Science-Projekts von immenser Wichtigkeit. Sie ermöglicht es Entscheidungsträger\_innen und Projektbeteiligten, schnell einen Überblick über die Stärken und Schwächen des Projekts zu erhalten und somit gezielt Bereiche für Verbesserungen zu identifizieren. Einfachheit und intuitive Interpretierbarkeit sind Schlüsselmerkmale, die eine effektive Kommunikation der Validierungsergebnisse ermöglichen und somit eine fundierte Entscheidungsfindung.