



---

### Highlights:

Wenn Sie den CRISP-DM-Standardprozess für Data Mining als Richtlinie nutzen, können Sie die Anforderungen an ein Data Mining-Tool für Ihr Unternehmen herausarbeiten.

---

## Die Evaluation eines Data Mining-Tools

### Schaffen Sie für Ihre Evaluation zunächst einen Rahmen

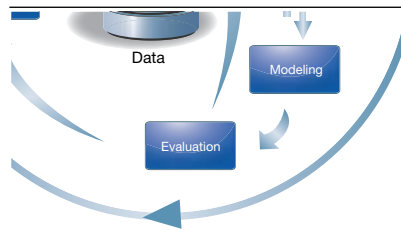
Mit Data Mining können Sie klügere Entscheidungen treffen, die zu wichtigen und konkreten Ergebnissen führen, z. B. erhöhten Umsätzen und effizienteren Prozessen. Zwar sind mit Data Mining äußerst vielversprechende Ergebnisse möglich, aber oftmals wird suggeriert, dass sich bei minimalem Aufwand unglaubliche Ergebnisse erzielen lassen. Wird mit dieser Erwartungshaltung ein Data Mining-Tool ausgewählt, kann der Return on Investment zu einer Enttäuschung werden.

Wenn Sie bei Ihrem Data Mining-Projekt bestimmte Ergebnisse erzielen möchten, bewerten Sie zunächst Ihre Geschäftssituation sowie die Leistung der Tools während des gesamten Data Mining-Prozesses. Als Unterstützung wurde mithilfe des anwendungsunabhängigen CRISP-DM-Standard für Data Mining (Cross-Industry Standard Process for Data Mining) die folgende Checkliste zusammengestellt.

Beim CRISP-DM wird der Data Mining-Prozess in sechs Phasen unterteilt (siehe Diagramm auf der nächsten Seite). Die Phasen müssen dabei nicht streng nacheinander durchlaufen werden. Vielmehr ist es notwendig, immer wieder zwischen den verschiedenen Phasen hin- und herzuspringen. Achten Sie bei der Evaluation Ihres Data Mining-Tools auf diesen interaktiven Aspekt. Das Tool, für das Sie sich entscheiden, sollte flexibel genug sein, um in jeder Projektphase Änderungen durchführen zu können.

Die folgenden Evaluationskriterien sind diesen sechs Phasen untergeordnet. Am Ende finden Sie einige zusätzliche Anmerkungen.





---

In Phase 1 legen Sie Ihre Projektziele fest und formulieren einen vorläufigen Plan für das Data Mining.

---

### CRISP-DM-Phase 1: Business Understanding

In der ersten Phase liegt der Schwerpunkt darauf, die Projektziele und -anforderungen aus betriebswirtschaftlicher Sicht zu verstehen. Anschließend muss anhand der dabei gewonnenen Erkenntnisse ein Data Mining-Problem definiert sowie ein vorläufiger Plan ausgearbeitet werden. Mit den Punkten auf der Checkliste soll untersucht werden, ob sich bei dem Tool die Erfolge, die bei einer großen Auswahl an Geschäftsproblemen verzeichnet werden konnten, protokollieren lassen und inwiefern das Tool das wirtschaftliche Verständnis mit der technischen Seite des Data Mining-Prozesses verbindet.

- Können mit dem Tool nachweislich viele verschiedene Geschäftsprobleme gelöst werden, einschließlich unserer individuellen Probleme?
  - War das Tool nachweislich bei der Lösung von Problemen hilfreich, die unsere Branche oder unsere Anwendung betreffen? Wurde das Tool bereits erfolgreich bei Anwendungen eingesetzt, die wir möglicherweise nutzen möchten?
- Inwiefern verknüpft das Tool das wirtschaftliche Verständnis mit der technischen Seite des Data Mining-Prozesses?
  - Können die Schritte, die bei der Verwendung des Tools durchgeführt werden, den Geschäftsanforderungen des Data Minings eindeutig zugeordnet werden? Werden dem Geschäftskunden Data Mining-Konzepte anschaulich präsentiert? Wie kann das Tool in Projektmanagement- oder andere Planungstools integriert werden? Müssen für das Tool Anwendungen geschrieben werden, um diese Lücke zu schließen?

---

In Phase 2 prüfen Sie, wie sich potenzielle Data Mining-Tools in Ihre derzeitige Technologieumgebung integrieren lassen, um so das Sammeln und Visualisieren von Daten zu ermöglichen.

---

## CRISP-DM-Phase 2: Datenverständnis

In dieser Phase werden zunächst Daten gesammelt. Im Anschluss wird sich im Rahmen verschiedener Maßnahmen mit diesen Daten auseinandergesetzt, um sich mit ihnen vertraut zu machen. Offene Standards für den Datenzugriff und Visualisierungstechniken sind wichtige Punkte, die in dieser Phase berücksichtigt werden sollten.

- Wie schützt das Tool meine bestehenden Investitionen in die IT-Infrastruktur?
  - Ist das Data Mining-Tool mit Ihren vorhandenen Datenspeichern kompatibel? Funktioniert das Tool mit offenen Datenstandards oder müssen Daten in proprietäre Dateiformate migriert werden?
- Inwiefern können mit dem Tool die Daten interaktiv untersucht und visualisiert werden?
  - Werden mit dem Data Mining-Tool Visualisierungstechniken zur Verfügung gestellt, mit denen sich Datenmuster erkennen lassen? Lässt sich die Visualisierung interaktiv durchführen, indem die Grafiken an sich verändert und neue Grafiken auf Basis der verschiedenen Datendimensionen erstellt werden können?

---

In Phase 3 untersuchen Sie, wie Daten für eine optimale Analyse von den verschiedenen Data Mining-Tools aufbereitet werden.

---

## CRISP-DM-Phase 3: Datenaufbereitung

In der Datenaufbereitungsphase werden sämtliche Aktivitäten abgedeckt, die der Erstellung des endgültigen Datensatzes aus den ursprünglichen Rohdaten dienen. Effizienz und Leichtigkeit der Datenaufbereitung gehören dabei zu den Punkten auf der Checkliste.

- Wie geht das Tool bei der Datenaufbereitung vor?
  - Wird der gesamte interaktive Data Mining-Prozess, einschließlich der Aktivitäten, die zur Aufbereitung der Rohdaten für die Modellerstellung benötigt werden, für ein effizientes Data Mining skaliert? Werden beim Tool die Schritte für die Datenaufbereitung anschaulich dargestellt?
- Können mit dem Tool die Daten für die Aufbereitung automatisch extrahiert werden?
  - Können mit dem Tool Daten automatisch extrahiert werden oder müssen SQL-Abfragen für Verknüpfungen, Aggregationen, Sortierungen und andere Datenaufbereitungsoperationen manuell geschrieben werden?

---

Die effektive Modellierung gehört zu den Schlüsselkomponenten des Data Minings. In Phase 4 ermitteln Sie, welche Data Mining-Tools mit den verschiedenen Modellierungstechniken am besten funktionieren.

---

## CRISP-DM-Phase 4: Modellierung

In dieser Phase werden verschiedene Modellierungstechniken ausgewählt und angewendet und Parameter optimal kalibriert. Analysten wechseln dabei häufig zurück in die Datenaufbereitungsphase, um den Anforderungen der verschiedenen Modelltypen zu entsprechen. Da ein bestimmtes Data Mining-Problem mit verschiedenen Modellen angegangen werden kann, beziehen sich die Punkte auf der Checkliste auch darauf, wie gut mit dem Tool verschiedene Techniken angewendet und miteinander verglichen werden können.

- Inwiefern erhöht das Tool die Produktivität von Analysten?
  - Können Analysten mit dem Tool schnell effektive Modelle entwickeln? Wie leicht können Benutzer verschiedene Modelle ausprobieren, um sich am Ende für die beste Lösung zu entscheiden? Wie leicht lässt sich die Datenaufbereitung bei bestimmten Modellanforderungen durchführen?
- Steht bei dem Tool eine große Auswahl an Techniken zur Verfügung?
  - Bietet das Tool Techniken oder Algorithmen zur Visualisierung, Klassifizierung, Assoziierung, Regression und zum Clustering?
- Können mit dem Tool Techniken miteinander kombiniert werden?
  - Können die verschiedenen Techniken problemlos kombiniert werden, um bessere Ergebnisse zu erhalten? Können die Ergebnisse der Algorithmen in den Datensatz zur Nachbearbeitung und Analyse eingegliedert werden?
- Schützt das Tool meine bestehende Technologieinvestitionen, wie Algorithmen und andere Tools?
  - Kann das Data Mining-Tool mit Ihren bestehenden Algorithmen arbeiten? Wie gut lässt sich das Data Mining-Tool gemeinsam mit Ihren anderen Datenanalysetools einsetzen?

---

In Phase 5, der Evaluationsphase, bewerten Sie die verschiedenen Datenmodelle, um zu gewährleisten, dass alle wirtschaftlichen Fragestellungen berücksichtigt wurden.

---

### CRISP-DM-Phase 5: Evaluation

Die Evaluationsphase besteht aus einer sorgfältigen Bewertung des Modells oder der Modelle, die vor der Bereitstellung verwendet wurden. Ein wichtiges Ziel hierbei ist es zu ermitteln, ob eine wichtige wirtschaftliche Fragestellung ausreichend berücksichtigt wurde. Die Punkte auf der Checkliste für die Evaluationsphase beziehen sich darauf, inwiefern die Interessen von Geschäftskunden beim Modell berücksichtigt wurden und wie gut die Ergebnisse für die entsprechende Zielgruppe aufbereitet wurden.

- Werden mit dem Tool durchgehend herausragende Ergebnisse erzielt?
  - Genügen die Ergebnisse, die mit Ihrem Data Mining-Tool generiert werden, stets hohen Standards oder werden nur in bestimmten Fällen oder bei bestimmten Datenarten gute Ergebnisse erzielt? Spiegeln Ihre Ergebnisse alle wirtschaftlichen Fragestellungen genau wider und ist deshalb die Leistung bei Testdaten überzeugend?
- Werden mit dem Tool Ergebnisse erzielt, die leicht verständlich sind?
  - Können Geschäftskunden Ihre Ergebnisse problemlos verstehen? Wenn dies nicht der Fall ist, welche Schritte sind für überzeugendere Ergebnisse notwendig? Wurden während des Data Mining-Prozesses Wirtschaftsexperten angeregt, beim Tool mitzuwirken?
- Kann die gesamte Bandbreite der Visualisierung angewendet werden, um die Ergebnisse eines Modells zu validieren?
  - Können Werte, Prognosen und andere Ergebnisse des Modells problemlos analysiert werden, um die Modellleistung zu validieren?

---

In Phase 6 der CRISP-DM-Checkliste ermitteln Sie, inwiefern die Data Mining-Lösungen problemlos und effizient bereitgestellt werden können.

---

### CRISP-DM-Phase 6: Bereitstellung

In der Bereitstellungsphase findet der Data Mining-Prozess Anwendung. Dabei kann durch den Prozess beispielsweise einfach ein Einblick in eine wirtschaftliche Fragestellung gewährt werden oder der Prozess wird in eine Anwendung implementiert, um aktuelle Erkenntnisse zu Informationszwecken zur Verfügung zu stellen. Bei der Bereitstellung sind häufig umfangreiche Services notwendig. Die Checkliste bezieht sich deshalb darauf, inwiefern dies mit dem Tool möglich ist.

- Wie kann ich meine Data Mining-Lösungen bereitstellen (heute und in der Zukunft)?
  - Wie können meine Data Mining-Lösungen in betriebliche Anwendungen integriert werden? Ist die Integration kostengünstig möglich oder erfordert sie erhebliche Entwicklungszeit und -kosten? Wie einfach können meine Lösungen aktualisiert werden? Wenn die Lösungen nicht ohne Weiteres aktualisiert werden können, welcher Aufwand und welche Kosten kommen auf mich zu?

### Zusätzliche Anmerkungen: Betriebskosten

Neben der Berücksichtigung der Data Mining-Phasen im Rahmen der CRISP-Vorgehensweise sollten Sie auch eine Analyse im Hinblick auf den Return on Investment durchführen.

- Wie gestalten sich die Betriebskosten?
  - Quantifizieren Sie die Betriebskosten über den gesamten Lebenszyklus des Produkts hinweg sowie das Serviceangebot, einschließlich notwendiger ergänzender Systeme. Quantifizieren Sie den zu erwartenden Umsatz. Wann kann mit einem positiven Return on Investment gerechnet werden und entspricht dies den Geschäftszielen?
- Wie stellt sich die Implementierungszeit dar?
  - Wie lange dauert die Implementierung des Data Mining-Tools? Werden andere Tools oder Hardware benötigt? In welchem Umfang sind Schulungen, Beratungen und Anpassungen notwendig, um beim Projekt Ergebnisse zu erzielen?
- Sind die Fähigkeiten der Benutzer (heute und in Zukunft) für die Verwendung des Tools ausreichend?
  - Welche Fähigkeiten sind notwendig, um mit dem Data Mining-Tool sinnvolle Ergebnisse zu erzielen? Ist das Tool für technisch sehr versierte Benutzer oder Data Mining-Anfänger ausgelegt? Oder kann es von beiden Gruppen bzw. von Benutzern mit durchschnittlichen Kenntnissen verwendet werden? Welche Schulungskosten werden entstehen, um jeden angemessen vorzubereiten? Denken Sie auch an die Fähigkeiten potenzieller Benutzer in der Zukunft.
- Kann das Tool für unsere Benutzer angepasst werden?
  - Wie kann das Tool für verschiedene Verwendungen angepasst werden? Können häufig durchgeführte Prozesse zur Wiederverwendung gespeichert werden? Können mit dem Tool Aufgaben automatisiert werden? Gibt es Services zum Anpassen von Benutzeroberflächen oder sonstige Unterstützung?

---

Wenn Sie die Effizienz eines Data Mining-Tools ermitteln, achten Sie auch auf die Gesamtbetriebskosten, den Return on Investment und die zukünftigen Möglichkeiten, die sich aus der Partnerschaft mit einem Hersteller ergeben.

---

### Zusätzliche Anmerkungen: Hersteller

Denken Sie bei der Bewertung Ihres Data Mining-Tools schließlich über den Einfluss des Herstellers nach. Der Kauf einer Data Mining-Lösung ist in vielerlei Hinsicht eine Investition in die Zukunft. Sie benötigen deshalb auf dem bevorstehenden Weg einen zuverlässigen Partner.

- Hat der Hersteller andere Tools und Services im Angebot, mit denen sich ähnliche Probleme lösen lassen?
  - Bietet der Hersteller weitere Data Mining- oder Datenanalysetools an? Stehen Beratung, Schulung, technische Unterstützung und andere Services zur Verfügung? Sind diese Services weltweit verfügbar?
- Ist auf den Hersteller im Hinblick auf Aktualisierungen für Software und Services Verlass und entspricht deren Bereitstellung meinen Bedürfnissen?
  - Gehört der Hersteller bei der Bereitstellung von Data Mining-Lösungen zu den führenden Anbietern? Verfügt der Hersteller über die notwendigen Ressourcen, um auch in Zukunft einen hochwertigen Service weiterhin anbieten zu können?

### Informationen zu IBM Business Analytics

Die Software IBM Business Analytics liefert umfassende, einheitliche und korrekte Informationen, denen Entscheidungsträger zum Verbessern der Unternehmensleistung vertrauen. Ein umfassendes Portfolio aus Geschäftsvorteilen, fortgeschrittener Analytik, finanziellen Vorteilen und Strategiemangement sowie Analyseanwendungen bietet Ihnen sofort klare und umsetzbare Einblicke in die aktuelle Leistung und gibt Ihnen die Möglichkeit, zukünftige Ergebnisse vorherzusagen.

Als Teil dieses Portfolios unterstützt IBM SPSS Predictive Analytics Software Organisationen, zukünftige Ereignisse vorherzusagen und proaktiv auf Basis dieser Erkenntnisse zu handeln, um bessere Geschäftsergebnisse zu erzielen. Kunden aus den Bereichen Wirtschaft, öffentlicher Verwaltung und Lehre verlassen sich weltweit auf IBM SPSS Technologie als Wettbewerbsvorteil zur Kundengewinnung, -bindung und Erhöhung der Kundenumsätze bei gleichzeitiger Betrugsreduzierung und Risikominimierung. Durch die Integration von IBM SPSS Software in ihre täglichen Prozesse werden Organisationen zur Predictive Enterprise – sie sind dadurch in der Lage Entscheidungen zu treffen und zu automatisieren, um die Geschäftsziele zu erreichen und einen messbaren Wettbewerbsvorteil zu gewinnen. Für mehr Informationen besuchen Sie bitte [www.ibm.com/spss/de](http://www.ibm.com/spss/de).



---

© Copyright IBM Corporation 2010

IBM Corporation  
Route 100  
Somers, NY 10589

Eingeschränkte Rechte für Mitarbeiter der US-Regierung – Benutzung, Duplizierung und Veröffentlichung beschränkt durch GSA ADP Schedule-Vertrag mit IBM Corp.

Hergestellt in den USA  
Mai 2010  
Alle Rechte vorbehalten

IBM, das IBM Logo, ibm.com, WebSphere, InfoSphere und Cognos sind Marken oder eingetragene Marken der International Business Machines Corporation in den USA und/oder anderen Ländern. Wenn diese oder andere eingetragene Markenbegriffe von IBM mit einem Markenzeichen (® oder TM) gekennzeichnet sind, wenn Sie zum ersten Mal in diesen Informationen vorkommen, weist dies darauf hin, dass es sich zum Zeitpunkt der Veröffentlichung dieser Informationen um in den USA eingetragene Marken oder Marken nach Gewohnheitsrecht von IBM handelt. Solche Marken können auch in anderen Ländern eingetragene Marken oder Marken nach Gewohnheitsrecht sein. Eine aktuelle Liste der Marken von IBM finden Sie im Internet unter „Copyright and trademark information“ ([www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)).

SPSS ist eine Marke von SPSS, Inc., an IBM Company, die in vielen Gerichtsbezirken weltweit eingetragen ist.

Weitere Unternehmens-, Produkt- und Servicenamen können Marken anderer Hersteller sein



Bereiten Sie bitte auf