



# DATA ANALYTICS

## Scoping Leitfaden - Infografik

	Descriptive	Diagnostic	Predictive	Prescriptive
<b>Fragestellung</b>	Was ist passiert?	Warum ist es passiert?	Was wird passieren?	Wie soll ich handeln?
<b>Ansatz</b>	Analyse und Visualisierung strukturierter Live Daten.	Ursachenanalyse auf Basis verwirrender Information.	Erkennung und Verwendung Muster, um bestimmte Ergebnisse mithilfe von Algorithmen vorherzusagen.	Fortschrittliche Analyse, um Handlungsempfehlungen auf Wahrscheinlichkeitsbasis abzugeben.
<b>Anwendung</b>	Ergebnisse von mehreren Bereichen Ihres Unternehmens zusammenfassen.	Fundiert herausfinden, was die Ursache eines Problems war.	Fundierte Vorhersage über wahrscheinliche Ergebnisse.	Unterstützung bei komplexen oder zeitkritischen Entscheidungen.
<b>Mehrwert</b>	Information	Zusammenhänge	Vorhersage	Empfehlung
<b>Zeitlicher Fokus</b>	Vergangenheit	Vergangenheit	Zukunft	Gegenwart & Zukunft
<b>Entwicklungs-Komplexität</b>	● ● ● ●	● ● ● ●	● ● ● ●	● ● ● ●
<b>Datenbedarf</b>	● ● ● ●	● ● ● ●	● ● ● ●	● ● ● ●
<b>Ressourcen-aufwand</b>	● ● ● ●	● ● ● ●	● ● ● ●	● ● ● ●
<b>Tools / Plattformen*</b>	R, Python, SQL, Matlab, Excel, SPSS, PowerBI, Qlik, Tableau, SAS, Splunk, Big Data**	R, Python, SQL, Matlab, Excel, SPSS, PowerBI, Qlik, Tableau, SAS, Splunk, Knime, Klipfolio, Big Data**	R, Python, Excel, SPSS, EViews, Stata, Knime, Rapidminer, Big Data**	R, Python, Anaconda, Stata, Knime, Google AI Plattform, Rapidminer, Big Data**
<b>Methoden</b>	<ul style="list-style-type: none"> <li>• Statistik</li> <li>• Data Aggregation</li> <li>• Dashboards</li> <li>• Business Intelligence</li> <li>• Clustering</li> <li>• KPIs &amp; Measures</li> </ul>	<ul style="list-style-type: none"> <li>• Data Mining</li> <li>• Business Intelligence</li> <li>• Factor Analysis</li> <li>• Root Cause Analysis</li> <li>• Clustering</li> <li>• Berichte</li> </ul>	<ul style="list-style-type: none"> <li>• Simulation</li> <li>• Machine Learning</li> <li>• Regression</li> <li>• Classification</li> <li>• Clustering</li> <li>• Zeitreihenanalyse</li> </ul>	<ul style="list-style-type: none"> <li>• (Meta-)Heuristics</li> <li>• Decision Trees</li> <li>• Machine Learning</li> <li>• Neural Networks</li> <li>• Classification</li> <li>• Deep Learning</li> </ul>
<b>Technische Rollen</b>	<ul style="list-style-type: none"> <li>• Data Engineer</li> <li>• Data Architekt</li> <li>• Business Analyst</li> <li>• Datenbankentwickler</li> </ul>	<ul style="list-style-type: none"> <li>• Data Engineer</li> <li>• Data Architekt</li> <li>• Business Analyst</li> </ul>	<ul style="list-style-type: none"> <li>• Statistiker</li> <li>• Data Scientist</li> <li>• Data Analyst</li> </ul>	<ul style="list-style-type: none"> <li>• Data Scientist</li> <li>• Data Analyst</li> <li>• KI Entwickler</li> </ul>
<b>Grenzen der Methoden</b>	<ul style="list-style-type: none"> <li>• Teildatenmenge der Vergangenheit</li> <li>• Bezug zum Business Kontext muss durch Experten hergestellt werden</li> </ul>	<ul style="list-style-type: none"> <li>• Aussagefähigkeit abhängig von der Datenqualität</li> <li>• Nutzen-Aufwand Verhältnis abhängig von der Problematik</li> </ul>	<ul style="list-style-type: none"> <li>• zukünftige Vermutung</li> <li>• keine Handlungsempfehlung</li> </ul>	<ul style="list-style-type: none"> <li>• Handlungsempfehlung nicht nachvollziehbar und nicht transparent</li> </ul>
<b>Beispiel</b>	Die Vertriebsanalyse eines Produktherstellers zeigt einen Absatzeinbruch für bestimmte Produkte in bestimmten Ländern.	Nach einem plötzlichen Volumenanstieg in der Notaufnahme kann ein Klinikum feststellen, dass die Symptome von Patienten -hohes Fieber, trockener Husten, Müdigkeit - auf denselben Infektionserreger hinweisen.	Ein Fitnessstudio kann Kunden identifizieren, die ihr Abo höchstwahrscheinlich nicht verlängern, und aufgrund historischer Daten Anreize vorschlagen, die sie zur Abo-Verlängerungen bringen könnten.	Das selbstfahrende Auto Waymo führt auf jeder Fahrt Millionen von Berechnungen durch, um zu entscheiden, wann und wo es abbiegen soll, ob es langsamer / schneller fährt und wann es die Spur wechselt.

\* Tools / Plattformen Quellen: [Gartner](#), [Capterra](#), [G2](#), [PredictiveAnalyticsToday](#)

\*\*indikative Big Data Tools: Flink, Hadoop, Hbase, Storm, MongoDB, Java, Scala, Spark, Samza aus Quellen: [1](#), [2](#), [3](#)



# DATA ANALYTICS

## Scoping Leitfaden - Glossar (Seite 1/2)

Begriff	Beschreibung
<u>Aggregation</u>	Das Zusammenstellen von Informationen aus Datenbanken mit der Absicht, kombinierte Datensätze für die Datenverarbeitung vorzubereiten.
<u>Algorithmus</u>	Eine eindeutige Handlungsvorschrift, die aus endlich vielen und wohldefinierten Einzelschritten besteht, zur Lösung eines Problems / einer Problemklasse
<u>Anaconda</u>	Eine Open-Source-Distribution für die Programmiersprachen Python und R mit Fokus auf Verarbeitung von großen Datenmengen, Vorhersageanalyse und wissenschaftliches Rechnen.
<u>Analyse</u>	Die systematische Untersuchung, bei der das untersuchte Objekt in seine Bestandteile (Elemente) zerlegt wird. Diese Elemente werden dabei auf der Grundlage von Kriterien erfasst und anschließend geordnet, untersucht und ausgewertet.
<u>Ansatz</u>	Descriptive Analytics befassen sich mit der Analyse vorhandenen Daten (z.B. Visualisierung Umsatz vom letzten Quartal). Diagnostic Analytics helfen bei der Ursachenfindung (Ausfallgrunde einer Anlage anhand Sensordaten). Predictive Analytics nutzen vorhandene Daten und versuchen Muster zu erkennen, um einen Blick in die Zukunft zu werfen (z.B. Verkaufsprognose in neuen Marktsegmenten). Prognostic Analytics erzeugen Empfehlungen über Was und Wann ein Ereignis passieren wird (z.B. Instandhaltung der Bohranlage in zwei Wochen, um Ausfall zu vermeiden).
<u>Anwendung</u>	Wobei hilft Ihnen der jeweilige Ansatz? Worauf zielt der Ansatz ab?
<u>Beispiele</u>	Praxisnahe Beispiele machen immer ein Thema greifbarer!
<u>Big Data</u>	Der Begriff bezeichnet Datenmengen, welche beispielsweise zu groß, zu komplex, zu schnelllebig oder zu schwach strukturiert sind, um sie mit manuellen und herkömmlichen Methoden der Datenverarbeitung auszuwerten.
<u>Business Analyst</u>	Er analysiert und berichtet historische Business-Daten. (Vergangenheitsorientiert)
<u>Business Intelligence</u>	Ist ein der Wirtschaftsinformatik zuzuordnender Begriff, der Verfahren und Prozesse zur systematischen Analyse des eigenen Unternehmens bezeichnet. Dies umfasst die Sammlung, Auswertung und Darstellung von Daten in elektronischer Form.
<u>Classification</u>	Das Problem der Identifizierung, zu welcher einer Reihe von Kategorien (Teilpopulationen) eine neue Beobachtung gehört, auf der Grundlage eines Trainingsatzes von Daten, die Beobachtungen (oder Instanzen) enthalten, deren Kategoriemitgliedschaft bekannt ist.
<u>Clustering</u>	Ein Verfahren zur Entdeckung von Ähnlichkeitsstrukturen in (meist relativ großen) Datenbeständen.
<u>Dashboard</u>	Ein Informationsmanagement-Tool, das Key Performance Indicators (KPI), Metriken und Schlüsseldatenpunkte visuell verfolgt, analysiert und anzeigt, um den Zustand eines Unternehmens, einer Abteilung oder eines bestimmten Prozesses zu überwachen.
<u>Data</u>	Eine Reihe von Werten qualitativer oder quantitativer Variablen über eine oder mehrere Personen oder Objekte.
<u>Data Analytics</u>	Eine Sammlung von Techniken zum Extrahieren von Wert aus Daten
<u>Data Analyst</u>	Er extrahiert Informationen aus externen und internen Datenbanken mithilfe verschiedener Methoden wie Datenbereinigung, Datenkonvertierung und Datenmodellierung. (Zukunftsorientiert)
<u>Data Architekt</u>	Er analysiert bestehende IT-Architekturen, berät auf technischer Ebene und verantwortet die Planung, Entwicklung und Bereitstellung neuer Data- und IT-Infrastrukturen von Unternehmen.
<u>Data Engineer</u>	Er entwickelt die Grundlagen für verschiedene Datenoperationen (Generierung, Speicherung, Pflege, Aufbereitung, Anreicherung und Weitergabe von Daten), an dem Data Scientists und Analyst arbeiten. (Vergangenheitsorientiert)
<u>Data Scientist</u>	Er versteht die Anforderungen des Unternehmens und analysiert Daten-Muster von der Vergangenheit, aufgrund deren er Prognosen für die Zukunft erstellt. Er verwendet Storytelling-Tools, um die Ergebnisse mit den Teammitgliedern zu kommunizieren. (Zukunftsorientiert)
<u>Datenbank-entwickler</u>	Er erarbeitet technische Datenbank-bezogene Konzepte (Indexe, Stored Procedures, Functions, Views und Trigger, passt das Datenbankschema je nach Modell - hierarchisch, relational, objektorientiert usw. - an) und entwickelt Datenbanksysteme (DBS).
<u>Datenbedarf</u>	Was wird an Datenmenge benötigt?
<u>Decision Trees</u>	Geordnete, gerichtete Bäume, die der Darstellung von Entscheidungsregeln dienen.
<u>Deep Learning</u>	Eine Methode des maschinellen Lernens, die künstliche neuronale Netze (KNN) mit zahlreichen Zwischenschichten (hidden layers) zwischen Eingabeschicht und Ausgabeschicht einsetzt und dadurch eine umfangreiche innere Struktur herausbildet.
<u>Entwicklungs-komplexität</u>	Wie komplex ist den Ansatz umzusetzen und zuverlässige / aussagekräftige Ergebnisse zu bekommen?
<u>EViews</u>	EViews (Kurzform für „Econometric Views“) ist eine kommerzielle, windows-basierte Ökonometrie-Software.
<u>Excel</u>	Das am weitesten verbreitete Tabellenkalkulationsprogramm.
<u>Factor Analysis</u>	Ein Verfahren der multivariaten Statistik. Es dient dazu, aus empirischen Beobachtungen vieler verschiedener manifester Variablen (Observablen, Statistische Variablen) auf wenige zugrunde liegende latente Variablen („Faktoren“) zu schließen.
<u>Fragestellung</u>	Der Absprungpunkt für die Entscheidung: Welche Kernfrage wird anhand von Data Analytics beantwortet?
<u>Google AI Plattform</u>	Google AI Plattform ermöglicht es Entwicklern, Data Scientists und Data Engineers im Bereich maschinelles Lernen ihre ML-Projekte von der Idee über die Produktion bis hin zur Bereitstellung schnell und kostengünstig umzusetzen. Mit Vorteilen wie Data Engineering und der Flexibilität ohne Lock-in unterstützt Sie die integrierte Toolkette der AI Plattform beim Erstellen und Ausführen eigener ML-Anwendungen.
<u>Grenzen der Methoden</u>	Welche Limitationen / Unzulänglichkeiten stellt jeder Ansatz dar?
<u>KI-Entwickler</u>	Er schreibt und entwickelt Algorithmen für Künstliche Intelligenz und analysiert und wertet KI-generierten Daten aus.



# DATA ANALYTICS

## Scoping Leitfaden - Glossar (Seite 2/2)

Begriff	Beschreibung
<u>Knime</u>	Eine freie Software für die interaktive Datenanalyse. KNIME ermöglicht durch das modulare Pipelining-Konzept die Integration zahlreicher Verfahren des maschinellen Lernens und des Data-Mining.
<u>KPIs</u>	Mit dem engl. Begriff Key Performance Indicator werden in der Betriebswirtschaftslehre allg. Kennzahlen bezeichnet, die sich auf den Erfolg, die Leistung oder Auslastung des Betriebs, seiner einzelnen organisatorischen Einheiten oder einer Maschine beziehen.
<u>Machine Learning</u>	Das Studium von Computeralgorithmen, die sich durch Erfahrung automatisch verbessern. Es wird als Teilmenge der künstlichen Intelligenz angesehen. Algorithmen für maschinelles Lernen bauen ein mathematisches Modell auf der Grundlage von Beispieldaten auf, die als "Trainingsdaten" bezeichnet werden, um Vorhersagen oder Entscheidungen zu treffen, ohne explizit dafür programmiert zu sein.
<u>Matlab</u>	Eine kommerzielle Software des US-amerikanischen Unternehmens MathWorks zur Lösung mathematischer Probleme und zur grafischen Darstellung der Ergebnisse. Matlab ist vor allem für numerische Berechnungen mithilfe von Matrizen ausgelegt, woher sich auch der Name ableitet: MATrix LABoratory.
<u>Measures</u>	Eine Maßzahl, die zur Quantifizierung dient und der eine Vorschrift zur quantitativen reproduzierbaren Messung einer Größe oder eines Zustandes oder Vorgangs zugrunde liegt.
Mehrwert	Was ist der vom Ansatz erzeugte Gewinn?
<u>(Meta-) Heuristics</u>	Jeder Ansatz zur Problemlösung oder Selbstfindung, bei dem eine praktische Methode angewendet wird, die nicht garantiert optimal, perfekt oder rational ist, aber dennoch ausreicht, um ein unmittelbares, kurzfristiges Ziel zu erreichen. Heuristics sind spezifisch und problemabhängig, Metaheuristics sind High-Level problemunabhängige algorithmische Rahmen, die eine Reihe von Richtlinien oder Strategien zur Entwicklung heuristischer Optimierungsalgorithmen bereitstellen.
Methoden	Welche fachlichen Methoden werden verwendet?
<u>Muster-erkennung</u>	Die Fähigkeit, in einer Menge von Daten Regelmäßigkeiten, Wiederholungen, Ähnlichkeiten oder Gesetzmäßigkeiten zu erkennen.
<u>Neural Networks</u>	Computersysteme, die von den biologischen neuronalen Netzen inspiriert sind, aus denen das Gehirn von Tieren besteht. Solche Systeme "lernen", Aufgaben anhand von Beispielen auszuführen, im Allgemeinen ohne mit aufgabenspezifischen Regeln programmiert zu sein.
<u>PowerBI</u>	Ein Geschäftsanalyse Dienst von Microsoft. Ziel ist es, interaktive Visualisierungen und Business-Intelligence-Funktionen mit einer Oberfläche bereitzustellen, die so einfach ist, dass Endbenutzer ihre eigenen Berichte und Dashboards erstellen können.
<u>Python</u>	Eine universelle, skalierbare, üblicherweise interpretierte, höhere Programmiersprache. Sie fördert einen gut lesbaren, knappen Programmierstil und ist essentiell für Data Science. Python bietet eine umfangreiche Datenbank mit Bibliotheken, künstlicher Intelligenz und maschinellem Lernen.
<u>Qlik</u>	Qlik ist ein Softwareunternehmen, welches die Business Intelligence-Werkzeuge QlikView und Qlik Sense als Hauptprodukte anbietet.
<u>R</u>	R ist eine freie Programmiersprache für statistische Berechnungen und Grafiken.
<u>Rapidminer</u>	Eine Umgebung für Maschinelles Lernen, Data-Mining, Text Mining, Web Mining, automatische Stimmungsanalyse aus Internet-Diskussionsforen (Sentiment Analysis, Opinion Mining), Zeitreihenanalyse und Prognose.
<u>Regression</u>	Bei der statistischen Modellierung ist die Regressionsanalyse eine Reihe statistischer Prozesse zum Schätzen der Beziehungen zwischen einer abhängigen Variablen und einer oder mehreren unabhängigen Variablen.
Ressourcen-aufwand	Was ist die Rolle des Menschen bei der Entscheidung? Welchen Input benötigt das Data Analytics Tool von den Mitarbeitern? Muss der Mensch aktiv eine Rolle spielen?
<u>Root Cause Analysis</u>	Beinhaltet die Erfassung von Fehlern, ihrer Ursachen und die statistische Auswertung dieser Daten, an die sich eine Bewertung und abgeleitete Maßnahmen zur Fehlerreduzierung (und Fehlerkostenreduzierung) anschließen.
<u>SAS</u>	Eine herstelleregebundene Softwareumgebung für die Auswertung von Daten, hauptsächlich von geschäftlichen Daten („Business Intelligence“), aber auch in der Forschung und in der öffentlichen Statistik, die eine der vollständigsten und bewährtesten Sammlungen von Statistikroutinen anbietet.
<u>Simulation</u>	Eine Vorgehensweise zur Analyse von Systemen, die für die theoretische oder formelmäßige Behandlung zu komplex sind. Bei der Simulation werden Experimente an einem Modell durchgeführt, um Erkenntnisse über das reale System zu gewinnen.
<u>Splunk</u>	Eine Log-, Monitoring- und Reporting-Plattform, die Daten nahezu jeder Art und aus nahezu jeder Quelle für Benutzer zugänglich und nutzbar macht.
<u>SPSS</u>	SPSS bezeichnet eine Marke der Softwarefirma IBM, unter welcher Statistik- und Analyse-Software entwickelt und vertrieben wird.
<u>SQL</u>	Eine Datenbanksprache zur Definition von Datenstrukturen in relationalen Datenbanken sowie zum Bearbeiten (Einfügen, Verändern, Löschen) und Abfragen von darauf basierenden Datenbeständen.
<u>Stata</u>	Ein universelles Statistik-Softwarepaket für Datenverwaltung, statistische Analyse, Grafiken, Simulationen, Regression und benutzerdefinierte Programmierung.
<u>Statistik</u>	Die Disziplin, die die Erfassung, Organisation, Analyse, Interpretation und Präsentation von Daten betrifft.
Statistiker	Er sammelt und analysiert qualitative/quantitative Daten („ein Data Scientist ist ein Statistiker, der mit der Technologie Schritt gehalten hat“).
<u>Tableau</u>	Tableau Software ist ein Hersteller von Visualisierungs-Software mit Schwerpunkt Datenvisualisierung und Reporting.
Technische Rollen	Welche(s) Kompetenzen / Team sind benötigt?
Tools / Plattformen	Welche sind die meist verwendeten Tools und Plattformen?
<u>Ursachen-analyse</u>	Beinhaltet die Erfassung von Fehlern, ihrer Ursachen und die statistische Auswertung dieser Daten, an die sich eine Bewertung und abgeleitete Maßnahmen zur Fehlerreduzierung (und Fehlerkostenreduzierung) anschließen.
<u>Vorhersage</u>	Eine Aussage über Ereignisse, Umweltzustände oder Entwicklung in der Zukunft.
Zeitlicher Fokus	Descriptive und Diagnostic Ansätze fokussieren auf der Vergangenheit. Predictive und Prognostic versuchen die Zukunft (evtl. die Gegenwart) vorherzusagen.
<u>Zeitreihen-analyse</u>	Die Disziplin, die sich mit der inferenzstatistischen Analyse von Zeitreihen und der Vorhersage (Trends) ihrer künftigen Entwicklung beschäftigt.