

Datenblatt

# Predictive Analytics

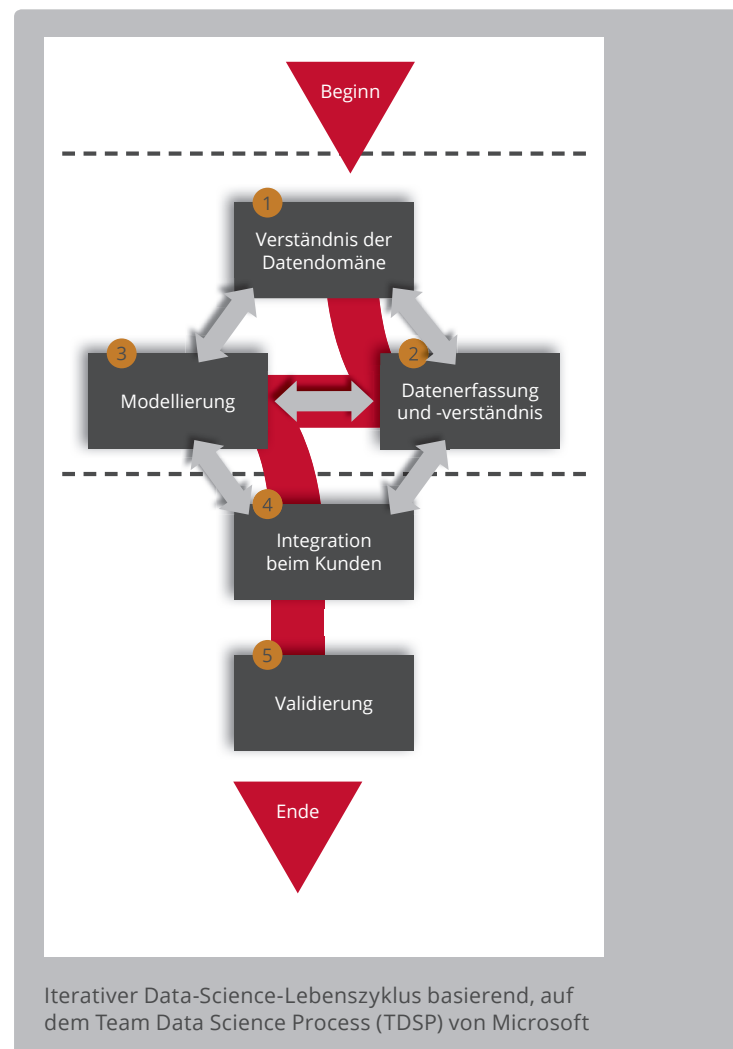
Kurz nach dem Start eines Formel-1-Rennens können die Teams ihre wahrscheinliche Endplatzierung vorhersagen und mit jeder Runde werden die Prognosen genauer. Dafür sammeln die Teams kontinuierlich große Datenmengen und analysieren sie mit mathematischen Modellen. Diese Modelle haben mittlerweile einen derart hohen Reifegrad erlangt, dass sie aus der Auswertung einer Vielzahl von Messpunkten präzise Vorhersagen ermöglichen. Auch in Unternehmen, Fertigungen und Produktionsstraßen fallen täglich große Datenmengen an, aus denen sich wertvolle Vorhersagen und bislang unbekanntes Wissen generieren lassen. Dieses Wissen bietet Unternehmen einen erheblichen Wettbewerbsvorteil.

Die infoteam Software AG ist ein etablierter Anbieter von Softwarelösungen und -dienstleistungen für Industrie, Life Science und Medizintechnik mit mehr als 200 Mitarbeitern. Für die Durchführung von Data-Analytics-Projekten orientieren wir uns an dem von Microsoft entwickelten Team Data Science Process (TDSP). Er ist sowohl mit Microsoft Azure Machine Learning nutzbar als auch als Leitfaden für Data-Science-Projekte verwendbar, die mit R, Python, InterSystems DeepSee® und iKnow oder anderen Tools durchgeführt werden.

Data-Science-Projekte lassen sich nicht linear abbilden, weshalb moderne agile Projektmanagementmethoden zum Einsatz kommen. Das iterative Vorgehen des TDSP harmonisiert hervorragend mit diesen Methoden. Neben einer standardisierten Projektstruktur liefert der TDSP auch eine Definition des Lebenszyklus von Data-Science-Projekten.

## Data-Science-Lebenszyklus:

Der Data-Science-Lebenszyklus gliedert sich in fünf Teilschritte:



### 1) Verständnis der Datendomäne

- Festlegung der Ziele und Anforderungen
- Aufbau eines Verständnisses des zugrunde liegenden Geschäftsmodells bzw. Produktionsprozesses
- Spezifikation und Verständnis der Datenquellen
- Identifikation der Schlüsselvariablen der Datendomäne

### 2) Datenerfassung und -verständnis

- Sammlung und erste Überprüfung der Daten
- Übernahme der Daten in infoteam-Standardstrukturen (optional)
- Prüfung der Datenqualität
- Vorbereitung des Datensatzes für die Modellierung
- Beurteilung und Konsolidierung
- Explorative Datenanalyse
- Konzeptionierung einer Datenpipeline (Datenfluss vom Sensor/Gerät etc. über die Datenbank bis hin zum Ergebnis aus Data-Science-Anwendungen)

### 3) Modellierung

- Vorauswahl der Modelle gemäß Kundenanforderungen und gefundenen Korrelationen
- Erzeugung neuer Variablen durch Transformation (Feature Engineering)
- Konstruktion und Evaluation mehrerer Modelle
- Auswahl und Optimierung eines Modells für den Einsatz im Produktionsumfeld
- Reality-Check: Abgleich des Modells mit dem Verständnis der Datendomäne

### 4) Integration beim Kunden

- Implementieren des Modells innerhalb der IT-Struktur des Kunden
- Einsatz des Modells und der Datenpipeline in einer Testumgebung
- Anschließender Einsatz des Modells und der Datenpipeline in der Produktivumgebung (Live-System)
- Erstellen eines Status-Dashboards zum Überwachen von Performance-Indikatoren des Modells

### 5) Validierung

- Finale Überprüfung der Datenpipeline und des Modells zusammen mit dem Kunden
- Review der Dokumentation
- Übergabe der Pipeline an die kundeneigenen IT-Experten oder Monitoring durch infoteam

Quelle: <https://docs.microsoft.com/de-de/azure/machine-learning/team-data-science-process/overview>

Alle verwendeten Hard- und Softwarenamen sind Handelsmarken und/oder eingetragene Marken der jeweiligen Hersteller.

© 2017, infoteam Software AG.  
Änderungen vorbehalten.

## Kontakt

Telefon: +49 (0) 9131 / 78 00 - 0  
Telefax: +49 (0) 9131 / 78 00 - 50  
info@infoteam.de

infoteam Software AG  
Am Bauhof 9  
D-91088 Bubenreuth