



RÜCKMELDUNGEN AUS DER DSAG ZU SAP PREDICTIVE ANALYTICS

Version 1.0

Walldorf, April 2018

Versionierung

Autor	Beschreibung
Dr. Ralf Finger	INFORMATION WORKS GmbH

Produktbezeichnungen

Im Interesse einer besseren Lesbarkeit des Leitfadens wird im Text in der Regel nur bei der ersten Verwendung der vollständige Produktname inklusive „SAP“ verwendet. Bei weiteren Verwendungen wird auf dieses Präfix verzichtet (z. B. „HANA“, statt „SAP HANA“).

Feedback

Feedback, Kommentare und konstruktive Kritik sind herzlich willkommen. Bitte posten Sie Ihre Beiträge direkt im DSAGNet unter <https://www.dsag.de/arbeitsgremien/ag-hana-analytics/artikel>

Mitwirkende

Die Bewertung von SAP Predictive Analytics wurde nur möglich, durch den aktiven Beitrag von Anwenderunternehmen, die im Rahmen von Expertengesprächen bereit waren, ihre Erfahrungen zu teilen – vielen Dank dafür. Die Gespräche waren auch die Grundlage für die Formulierung der Bewertungs-Statements in Abschnitt 4. Die Ansprechpartner von Anwenderunternehmen werden aus Anonymitätsgründen nicht persönlich genannt. Dies gilt auch für zwei Anwendungssituationen die unter Wahrung der Anonymität dankenswerter Weise von **Camelot ITLab GmbH** beigesteuert wurden.

Sprecherteam der AG HANA Analytics

Name	Firma
Gesa Fuchs	Ferrero Deutschland GmbH
Dr. Ralf Finger	INFORMATION WORKS GmbH

INHALT

1. MANAGEMENT SUMMARY / KERNAUSSAGEN	4
2. EINFÜHRUNG	7
2.1. Motivation und Zielsetzung	7
2.2. Vorgehensweise und Aufbau des Dokuments	7
3. PREDICTIVE-ANALYTICS-ANFORDERUNGEN.....	8
3.1. Cross Industry Standard Process for Data Mining (CRISP-DM).....	8
3.2. Datenexploration / Business and Data Understanding	9
3.3. Datenvorbereitung / Data Preparation	9
3.4. Modellierung / Modeling	9
3.5. Bewertung / Evaluation	10
3.6. Bereitstellung / Deployment	10
4. SAP PREDICTIVE ANALYTICS – RÜCKMELDUNGEN DER ANWENDER.....	11
4.1. Überblick	11
4.2. Desktop-Edition.....	12
4.3. Enterprise-Edition.....	15
4.4. R-Integration	18
5. AUSBLICK.....	20
6. ANHANG.....	21
7. IMPRESSUM.....	23

ABBILDUNGSVERZEICHNIS

Abbildung 1: SAP Predictive Analytics - Predictive Factory	4
Abbildung 2: Cross Industry Standard Process for Data Mining (CRISP) im Überblick	8
Abbildung 3: Die SAP Predictive Analytics Architektur im Überblick.....	11
Abbildung 4: SAP Predictive Analytics - Desktop Edition.....	12
Abbildung 5: SAP Predictive Analytics - Enterprise Edition	15
Abbildung 6: Integration von R und SAP HANA.....	18

1. MANAGEMENT SUMMARY / KERNAUSSAGEN

Im letzten Jahrzehnt haben Analysewerkzeuge zur Auswertung großer Datenmengen erhebliche Bedeutung gewonnen, da man Daten leichter denn je zuvor zusammentragen kann. Die Fähigkeit, Daten zusammenzutragen und zu analysieren, ermöglicht Unternehmen u.a. einen schnelleren, effizienteren und verlässlichen Entscheidungsfindungsprozess.

Im Zuge der Digitalisierung sind prädiktive Analyseinstrumente ein Weg, den Firmen handfeste Wettbewerbsvorteile zu liefern. Prädiktive Algorithmen mit entsprechenden Wahrscheinlichkeiten helfen, zukünftige Ereignisse einzuschätzen und die Unternehmen entsprechend zu steuern. So werden Big-Data-Themen in geschäftlichen Mehrwert umgewandelt.

SAP Predictive Analytics ist in unterschiedlichen Editionen erhältlich, die in diesem Dokument kurz vorgestellt werden [SAP20171026]. Diese reichen von rein Desktop-basierten Installationen bis hin zu einer integrierten Predictive Factory (vgl. Abbildung 1).

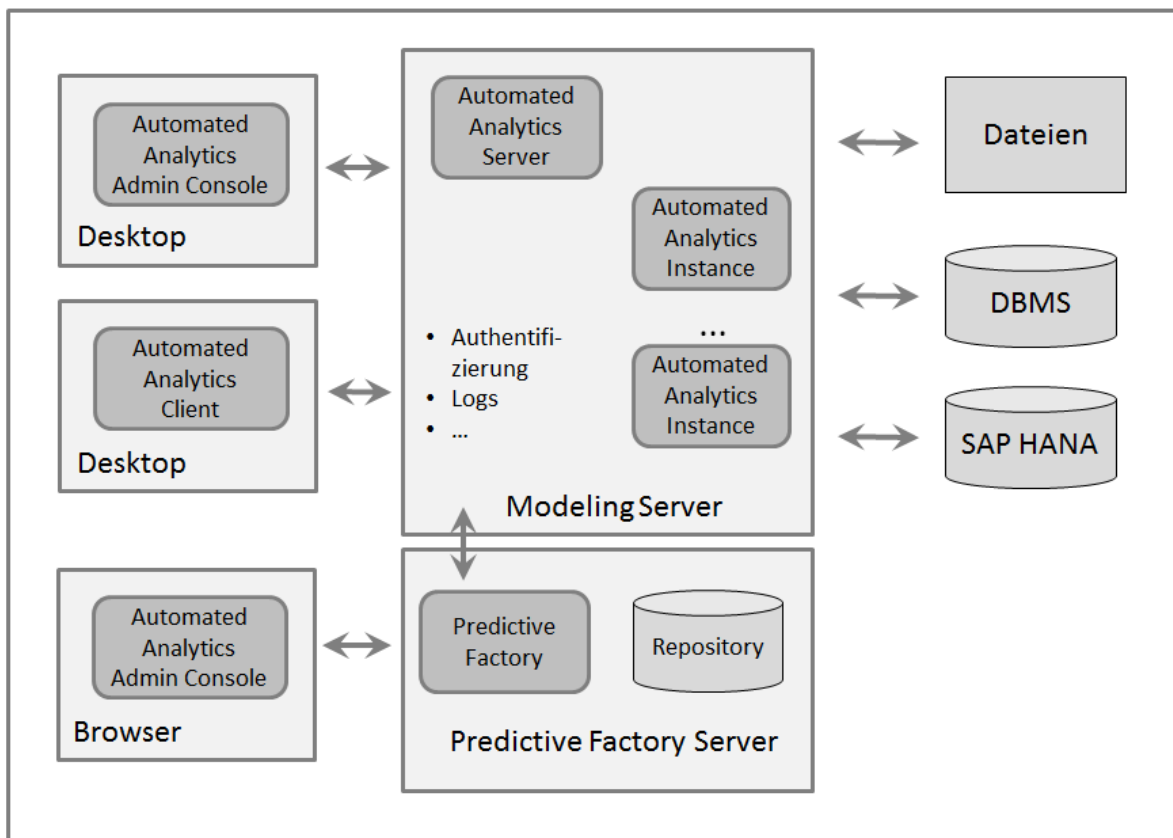


Abbildung 1: SAP Predictive Analytics – Predictive Factory

Im Rahmen der hier dargestellten Erhebung wurden zehn Anwenderunternehmen befragt. Diese nutzen SAP Predictive Analytics in unterschiedlichem Umfang. Einige gaben ihre Erfahrungen aus Proof-of-Concepts an, andere vor dem Hintergrund des Produktiveinsatzes. Keines der befragten Anwenderunternehmen nutzte jedoch SAP Predictive Factory. Diese hätte möglicherweise Teile der Kritikpunkte adressieren können. So ergänzt die SAP Predictive Factory die Enterprise Edition durch die Fähigkeit, Analyseprozesse und Modelle standardisiert abzulegen und zu verwalten. Dazu gehören auch das automatische Retraining und die Anwendung der Modelle auf neue Daten. Hierfür wird ein Repository implementiert, das mit dem Modeling Server verknüpft ist. Außerdem ist das Produkt in der Handhabung und im Zusammenspiel mit SAP HANA deutlich integriert.

Feststellen lässt sich, dass aufgrund der Erfahrungen im Rahmen der Proof-of-Concepts kein Produktiveinsatz der Software befürwortet wurde.

Die Rückmeldungen zum Produkt SAP Predictive Analytics sind vor dem Hintergrund der eingesetzten Editionen überwiegend eher kritisch. Und dies obgleich SAP und hier mit SAP HANA das Potenzial zugesprochen wird, Methoden sowie Werkzeuge zur prediktiven Analyse großer Datenmengen rasch zu adaptieren und für Unternehmen verfügbar machen zu können.

Die prädiktiven Ergebnisse lassen den Schluss zu, dass die prädiktive Analytik im Automatic Mode von Predictive Analytics in der Lage ist, genauso viele richtige Vorhersagen und Fehlklassifizierungen zu liefern, wie die in die Tiefe gehenden Ansätze einer auf „händischer“ R-Programmierung basierten Analyse. Der Automatic Mode bringt somit in Anwendungsfall-Szenarien der Unternehmen eine hervorragende Gesamtleistung. Dieses Merkmal unterstreicht auch die Verlässlichkeit der zugrunde liegenden Algorithmen.

Grundsätzlicher Kritikpunkt ist das technisch wenig ausgereifte Gesamtbild der Anwendung.

Dieser Gesamteindruck nährt sich z. B. aus folgenden Aspekten:

- Uneinheitliches Bild in der Frontend-Strategie SAP BO und UI5-Neuentwicklung. Hier besteht der klare Wunsch der Anwender, die Integration in eine einheitliche Frontend-Welt voranzutreiben. Hier ist die Strategie der SAP, alle Fähigkeiten Browser-basiert anzubieten (UI5 / Fiori „Look-and-Feel“) und so die Desktop-Umgebung abzulösen. Solange dies nicht erfolgt ist, müssen Kunden in eigene UI5-Entwicklung investieren.
- Heterogene Infrastruktur auf Backend-Seite: Die Predictive Factory verlangte etwa einen Windows Server, während andere Komponenten (z. B. R-Integration) Linux benötigen. Die Vielfalt der Systemvoraussetzungen unterschiedlicher Komponenten ist für ein Rechenzentrum kaum handhabbar. (Dies hat SAP inzwischen adressiert: Die SAP Predictive Factory ist ab der Version 3.3 auch auf Linux lauffähig.)
- Insgesamt muss die Plattform viel enger und nahtloser mit der HANA-Datenbank integriert werden. Ein flexibles Arbeiten aus den Frontends und eine einfache Konfiguration im Backend fehlen. Ausnahme bildet hier die Integration der APL, welche eine gute Integration mit SAP PA bietet.

- Der Gesamteindruck ist, dass man die KXEN als zugekauftes Produkt mit einfachsten Mitteln gekapselt hat, ohne dieses wirklich zu integrieren. SAP PA ist eher eine Hülle um den KXEN Kernel. SAP PA wird derzeit stark von der IT-Organisation in den Anwenderunternehmen getrieben, da hier ein Interesse an einem Ausbau der SAP-Plattform besteht. Eine saubere Integration mit der SAP-Plattform ist allerdings nicht gewährleistet, so dass der Mehrwert einer Integration nicht dargelegt werden kann.
- Des Weiteren sind die Werkzeuge für eine Arbeit im Fachbereich für Data Scientists nicht flexibel genug. Aus deren Sicht muss sich SAP PA mit bereits etablierten Open Source und kommerziellen Produkten messen lassen bzw. mit diesen integrieren. Sowohl in der Integrationsfähigkeit als auch in der Funktionalität ist SAP PA hier jedoch noch unterlegen. Dabei existieren bei Anwenderunternehmen jedoch auch Stimmen, die einen möglichen Anspruch der SAP, eine überzeugende Data Scientists Workbench zu entwickeln, skeptisch betrachten. Mitunter wird angeführt, dass dies von Wettbewerbsangeboten bereits gut abgedeckt wird. Empfehlung und Anwenderwunsch sind hier, den Automated Mode als Alleinstellungsmerkmal zu stärken. Diese Forderung deckt sich auch mit dem Eigenanspruch der SAP. So wird seitens der SAP stets betont, dass SAP PA die „annahmefreie“ statistische Datenanalyse unterstützen soll und dass der Expert Mode eher als Add-on zu verstehen ist. Vor dem Hintergrund dieses Anspruches erübrigen sich theoretisch die klassischen Anforderungen an Data Science Workbenches.
- Für eine Arbeit in den Fachbereichen durch eher unerfahrene Anwender (im Automated Mode) können (und müssen) viele Einstellungen vorgenommen werden, die unmittelbaren Einfluss auf die prädiktiven Ergebnisse haben und somit die Entscheidungsfindung bzw. Bewertung der Daten erheblich (und somit auch nachteilig) beeinflussen können. Die Schnittstelle Mensch-Maschine sollte dahingehend verbessert werden, dass z. B. angezeigt wird, welchen kritischen Einfluss falsche Eingaben haben, damit der Benutzer korrekt abschätzen kann, wie präzise das Modell arbeitet, und damit z. B. Fehlklassifizierungen vorgebeugt wird.
- Schließlich kritisieren Anwenderunternehmen auch die wenig differenzierten Schulungsmöglichkeiten als auch die begrenzte Dokumentation. Wettbewerber sind hier, auch aufgrund ihrer längeren Marktpräsenz, deutlich schlagkräftiger aufgestellt. Der interessierte Leser sei auf die im Anhang zusammengestellte Dokumentationssammlung hingewiesen.

Es besteht bei den Unternehmen daher ein nachvollziehbarer Wunsch der Bereitstellung eines SAP integrierten Produkts zur prädiktiven Analyse ihrer Datenbestände. Hier muss SAP „liefern“, will sie dieses Trendthema nicht verpassen. Mit der Abstellung der Kritikpunkte der Unternehmen geht gleichzeitig der Wunsch einher, die Weiterentwicklung des Produkts auch hin zu einer präskriptiven Anwendung gemeinsam mit der AG HANA Analytics durchzuführen.

2. EINFÜHRUNG

2.1. Motivation und Zielsetzung

SAP hat mit SAP Predictive Analytics ein Software Werkzeug mit dem Anspruch auf den Markt gebracht, Aufgabenstellungen der Predictive Analytics umfassend zu unterstützen. Die Erfahrungen von Anwenderunternehmen mit diesem Werkzeugangebot sind jedoch sehr unterschiedlich. Einerseits wird typischerweise der Ansatz des Automated Statistical Learnings gelobt. Andererseits wird nicht selten in punkto Usability und technischer Integration Kritik geübt.

Zielsetzung dieses Dokuments ist es, die Erfahrungen von Anwenderunternehmen zusammenzufassen, die SAP Predictive Analytics im praktischen Einsatz haben oder aber die Software im Rahmen von Teststellungen oder Proof-of-Concepts untersucht haben. Durch die strukturierte Aufnahme und Aufbereitung dieser Erfahrungen soll die Kommunikation zwischen den in der DSAG repräsentierten SAP-Anwenderunternehmen und der SAP als Softwareanbieter gefördert werden, um SAP Predictive Analytics weiterzuentwickeln.

2.2. Vorgehensweise und Aufbau des Dokuments

Ausgangspunkt der Betrachtung von SAP Predictive Analytics bilden die Anforderungen an Predictive Analytics. Diese werden in Kapitel 3 im Überblick dargestellt. Die Grundlage hierfür bildet der Cross Industry Process for Data Mining (CRISP-DM), der als wohletablierter Vorgehensstandard gelten kann.

Im Anschluss daran werden in Kapitel 4 die wichtigsten Komponenten von SAP Predictive Analytics vorgestellt. Auf dieser Basis sind dann die Aussagen der Anwenderunternehmen zugeordnet, die dabei als Positiv- und Negativaussagen unterschieden werden. Diese Aussagen wurden durch strukturierte Interviews mit den beteiligten Anwenderunternehmen erhoben und abgestimmt.

Alle Interviewpartner haben um Anonymität gebeten. Auch wurde den Interviewpartnern freigestellt, Fragenbereiche auszuklammern, wenn z. B. dadurch wettbewerbsrelevante Informationen tangiert gewesen wären. Um die Anonymität zu wahren, wird auf eine fallbezogene integrierte Darstellung der erhobenen Sachverhalte verzichtet, aus der ein Rückschluss auf das Anwenderunternehmen möglich wäre.

Basis der Auswertung bildet das Feedback von zehn Unternehmen. Je vier der erhobenen Situationen sind zum Zeitpunkt der Ergebnisdokumentation produktiv, in vier Situationen wurde ein Proof-of-Concept durchgeführt, in einer Situation wurde eine produktive SAP-PA-Installation durch ein Open-Source-Produkt abgelöst.

Von den zehn Interviews basieren sechs Fälle auf der Version 3, drei Fälle auf den Versionen 2.3, 2.3.1, 2.4 und ein Fall auf den Erfahrungswerten der Versionen 2 und 3.

3. PREDICTIVE-ANALYTICS-ANFORDERUNGEN

3.1. Cross Industry Standard Process for Data Mining (CRISP-DM)

Der Cross Industry Standard Process for Data Mining (CRISP-DM) ist heute der De-Facto-Standard zur Beschreibung des Vorgehens im Rahmen von Predictive Analytics Projekten (vgl. Abbildung 2).

Zentrale Anwenderrolle im CRISP-DM ist der Data Scientist. Hierunter wird typischerweise ein hochqualifizierter Analyst verstanden, der in einer Person Fähigkeiten der statistischen Datenanalyse, fachliches Wissen und Datenmanagement-Kenntnisse vereint. Diese Anwenderrolle führt maßgeblich die genannten Tätigkeiten aus.

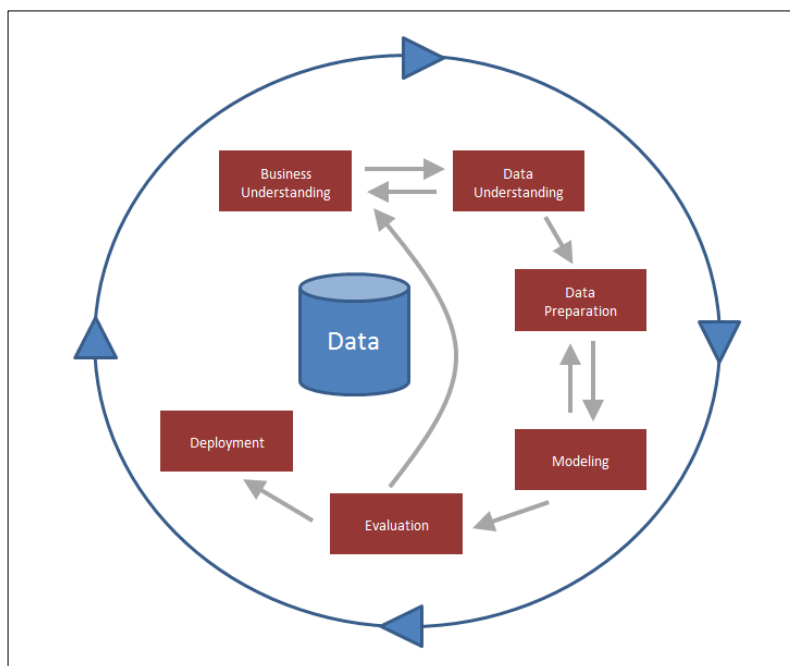


Abbildung 2: Cross Industry Standard Process for Data Mining (CRISP) im Überblick

Das Vorgehen des CRISP-DM zeigt den für Projekte des Predictive Analytics typischen, zyklischen Aufbau. Darüber hinaus sind insbesondere in den Aktivitäten der Anforderungsaufnahme und der explorativen Datenanalyse sowie zwischen den Aktivitäten der Datenvorbereitung und der Modellbildung ausgeprägte Iterationen üblich.

Diese Charakteristik in der Arbeitsweise ist maßgeblich durch Software-Werkzeuge zu unterstützen. Besonders hervorzuheben ist dabei, dass der Wechsel zwischen diesen Tätigkeiten sehr interaktiv erfolgen muss, d.h. unterstützende Werkzeuge müssen hochgradig integriert sein. Des Weiteren werden im Rahmen von Predictive-Analytics-Projekten i. d. R. große Datenmengen verarbeitet. Dies macht es erforderlich, dass Verarbeitungsprozesse problemlos Server-basiert und In-Database ausgelöst werden können.

3.2. Datenexploration / Business and Data Understanding

Zu Beginn eines Predictive-Analytics-Projekts ist die geschäftliche Fragestellung aufzunehmen und im Hinblick auf die notwendige Datengrundlage zu untersuchen. Aus diesem Grund ist in dieser Phase eine enge Zusammenarbeit zwischen Data Scientist und weiteren fachlichen Domänen-Experten anzustreben.

Diese Zusammenarbeit ist zum einen stark auf die interaktive Visualisierung des Datenmaterials angewiesen. Dazu gehört die Fähigkeit, flexibel und automatisiert grafische Auswertungssichten (Streudiagramme, Verteilungen etc.) zu erzeugen. Neben der Aufdeckung erster Muster in den Daten wird dabei typischerweise auch die Datenqualität untersucht.

Darüber hinaus müssen i. d. R. unterschiedliche Datenquellen schnell kombiniert werden. Diese Datenquellen können sehr unterschiedlicher Natur sein und von relationalen Strukturen bis zu Big Data reichen.

3.3. Datenvorbereitung / Data Preparation

Die Datenvorbereitung dient der Aufbereitung des grundlegenden Datenmaterials für die Modellbildung. Dies umfasst alle denkbaren Arten von Wertableitungen. Dazu gehören Filterung, Überleitungen von Feldinhalten, Aggregationen, Anwendung statistischer Funktionen etc.

3.4. Modellierung / Modeling

Die Modellierung ist der eigentliche Kern eines Predictive-Analytics-Projekts. Historische Daten mit bekannten Ergebnissen werden genutzt, um mittels statistischer Verfahren Vorhersagemodelle zu errechnen. Letztere erlauben dann die Vorhersage interessierender Ereignisse in neuen Daten.

Dazu werden die Auswertungsdaten in Zielvariablen und Prädiktoren unterteilt. Auf diese werden alternative statistische Verfahren zur Berechnung eines Vorhersagemodell angewandt (sog. Training). Hierbei ist die Breite und Differenzierung der unterstützten Modelle ein wichtiger Faktor (Entscheidungsbäume, Cluster-Analysen, Neuronale Netze etc.).

Ergebnis der Modellbildung ist ein Regelwerk, das nach Maßgabe der statistischen Modelle interessierende Ereignisse in neuen Daten bewerten kann. Typische Beispiele hierfür sind Kundenreaktionen auf Kampagnen, Forderungsausfälle oder der Ausfall von Maschinen im Fertigungsprozess.

3.5. Bewertung / Evaluation

Charakteristisch für Predictive-Analytics-Projekte ist die automatisierte Berechnung zahlreicher Modelle. Zielsetzung der Bewertung ist der Vergleich der errechneten Modelle hinsichtlich ihrer Prognosegüte. Dies kann auf Basis der für die Modellbildung genutzten historischen Daten erfolgen. Ein alternatives Vorgehen besteht darin, die Modellbildung nur auf einem Teil der historischen Daten durchzuführen (z. B. durch Stichprobenziehung) und das errechnete Modell auf die übrigen historischen Daten anzuwenden, für die das tatsächliche Ergebnis ebenfalls bekannt ist.

3.6. Bereitstellung / Deployment

Die Bereitstellung bezeichnet die Überführung der errechneten Vorhersagemodelle in die operative Nutzung. Hierbei werden die errechneten Modelle in definierten Zeitintervallen oder in Echtzeit auf neue Daten angewandt. Die Ergebnisse dieser Vorhersageberechnungen können dann als Bericht oder Datei bereitgestellt oder in Echtzeit an Geschäftsprozesse weitergeben werden.

Darüber hinaus ist das Retraining eine wichtige Fähigkeit einer Plattform für das Deployment von Predictive-Analytics-Modellen. Hierbei werden die vorliegenden Entscheidungsmodelle mit aktualisierten historischen Daten erneut ermittelt, um diese vor dem Hintergrund des neuen Datenstandes nach zu schärfen. Dabei kann es notwendig sein, dieses Retraining in regelmäßigen Abständen zu wiederholen sowie alternative Modelle im Betrieb auf ihre Prognosegüte hin regelmäßig zu vergleichen.

4. SAP PREDICTIVE ANALYTICS – RÜCKMELDUNGEN DER ANWENDER

4.1. Überblick

SAP Predictive Analytics besteht aus verschiedenen Werkzeugen. Oft wird von Anwenderunternehmen hierbei insbesondere deren Integration in sich als auch mit dem SAP-Portfolio insgesamt gefordert. Dies wird seitens SAP angestrebt und typischerweise ebenso positioniert (vgl. Abbildung 3).

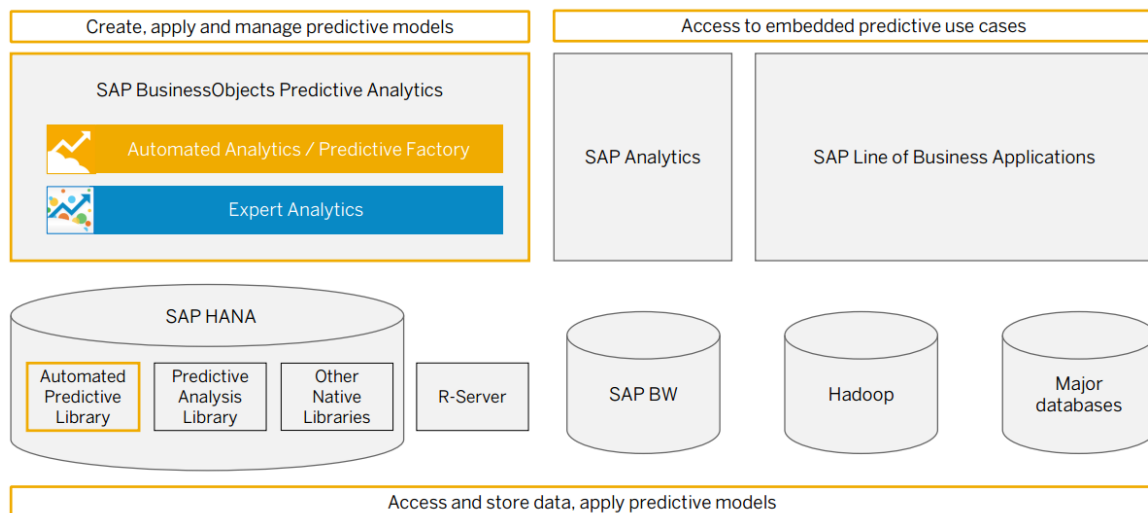


Abbildung 3: Die SAP-Predictive-Analytics-Architektur im Überblick

In der Praxis sind aktuell verschiedene sogenannte „Editionen“ von SAP Predictive Analytics zu unterscheiden, die ein unterschiedlich breites Leistungsspektrum aufweisen. Deren Integration mit der SAP Plattform insgesamt muss i. d. R. heute noch durch ein Systemintegrationsprojekt hergestellt werden.

Nachfolgend werden diese Editionen im Einzelnen vorgestellt und die Rückmeldungen der Anwender zugeordnet.

4.2. Desktop-Edition

Die Einstiegs-Edition von SAP Predictive Analytics basiert auf einer lokalen Installation auf einem Desktop. Diese Installation besteht aus zwei Clients, die unter einer gemeinsamen Oberfläche für deren Aufruf zusammengeführt sind. Im Einzelnen:

- Im **Expert Mode** modelliert der Data Scientist seine Datenanalysen mit sehr hohen Freiheitsgraden. Dies setzt beim Bediener umfassende Kenntnisse des Datenmanagements und der statistischen Modellierung voraus.
- Der **Automated Mode** erlaubt hingegen die ausgeprägt automatisierte Generierung statistischer Modelle. Grundlage hierfür ist lediglich die Definition von Datenfeldern als Prädiktoren und die Benennung des zu erklärenden Ereignisses in den Daten. Die Auswahl geeigneter Verfahren, deren Berechnung und die Bewertung der Prognosegüte erfolgt dann vollautomatisch, wobei statistische Details in Verarbeitungs-Logs einsehbar sind.

Diese Verarbeitungstypen können auf relationale Datenbanken, Dateien und SAP HANA zugreifen (vgl. Abbildung 4). Dabei werden allerdings die zu verarbeitenden Daten für die Berechnungen an die Desktop-Anwendung übergeben, sodass die Datenmenge für diese Art der Verarbeitung begrenzt ist.

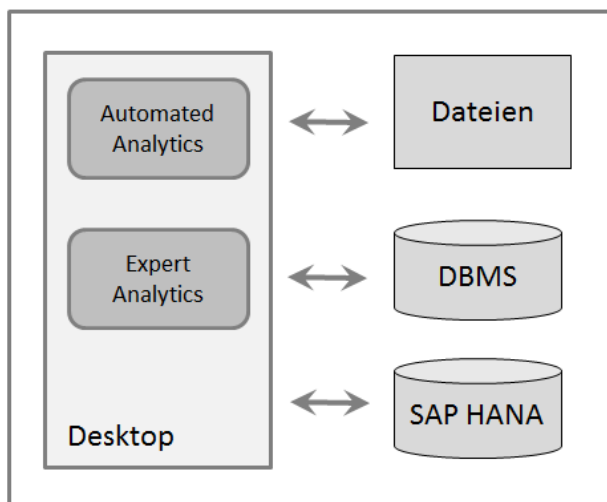


Abbildung 4: SAP Predictive Analytics – Desktop-Edition

Die folgenden Erfahrungen wurden erhoben:

CRISP-DM	Bewertung
Datenexploration	<ul style="list-style-type: none"> - Die Datenexploration im Expert Mode wird als nicht flexibel genug angesehen - Daten müssen für Visualisierungen oft nach Excel exportiert werden, um dort Charts zu erzeugen. Die vorhandenen Oberflächen leisten nicht das, was man im Projekt braucht - Es existieren harte Fehler - Wichtiger Kritikpunkt ist die wenig integrierte Frontend-Welt. Die Angebote rund um SAP BO, UI5 und HANA-Integration wirken hier inkonsistent
Datenvorbereitung	<ul style="list-style-type: none"> - Datenvorbereitungen können nicht effizient mit dem Data Manager durchgeführt werden. Dadurch ist man gezwungen, HANA Scripting oder SAP BW zu nutzen. Beides sind aber Umgebungen, die eher IT-orientiert sind und für Fachbereiche nicht geeignet - Verarbeitungen sind nicht leicht wiederholbar, gewünscht wäre hier auch eine Collaboration-orientierte Datenablage - Fachbereichsanwender, die Datenvorbereitungen im Expert Mode durchführen, empfinden die Arbeit als umständlich - Kernansprechpartner kommen „unter Schmerzen“ zurecht. Typische Fachanwender können aber nicht überzeugt werden
Modellierung	<ul style="list-style-type: none"> + Der Automated Mode wird generell als hilfreich empfunden + Mit dem Automated Mode werden auch die Prognose-Ergebnisse anderer Analytics-Werkzeuge gegengerechnet. Die Modellgüte wird positiv bewertet + Mitunter gibt es auch nur Verdachtsmomente für einen Zusammenhang. Durch den Automated Mode kann man schnell erste Aussagen automatisch generieren, was als hilfreich empfunden wird + Positiv wurde bewertet, dass SAP PA über ausgewählte Verfahren zur Modellbildung verfügt, die in anderen Werkzeugen nicht ohne weiteres zur Verfügung stehen (z. B. Ridge Regression) - Auch bei einfachen Sachverhalten wird schnell ein custom R-Knoten im Expert Mode erforderlich - Die Berichte im Automated Mode wirken optisch nicht sehr modern - Ergebnisvisualisierungen im Automated Mode enthalten zu wenig für den Statistiker, um das Ergebnis wirklich zu interpretieren. Die Interaktionswerte helfen nicht. Stattdessen werden erprobte statistische Signifikanzen und andere Kennwerte benötigt. Die gezeigten Grafiken zur Wichtigkeit einer Variablen sind in der SAP-PA-Dokumentation nicht wirklich beschrieben, daher ist die Interpretation nicht klar

Bewertung	<ul style="list-style-type: none"> + Der automatische Modellvergleich wird grundsätzlich als hilfreich angesehen - Wettbewerbsprodukte stellen die automatische Modellbildung und Bewertung intuitiv zur Verfügung - Die automatische Modellbildung ist auf die hinterlegten Verfahren begrenzt
Bereitstellung	<i>Für Desktop-Edition nicht relevant</i>

+ positive Bewertung
- negative Bewertung

4.3. Enterprise-Edition

Die Enterprise-Edition adressiert das Problem der Verarbeitung großer Datenmengen. So wird mit dieser Edition ein Modeling Server etabliert, durch den die Clients rechenintensive Prozesse an Server-Prozesse abgeben können. Hierbei können auch mehrere Server-Instanzen gleichzeitig gestartet werden, um parallele Verarbeitungen zu unterstützen (vgl. Abbildung 5).

Ergänzt wird diese Infrastruktur durch eine Administrationskonsole, um diese Modellierungsserver zu verwalten einschließlich z. B. der Authentifizierungsregeln für den Zugriff auf diese Server.

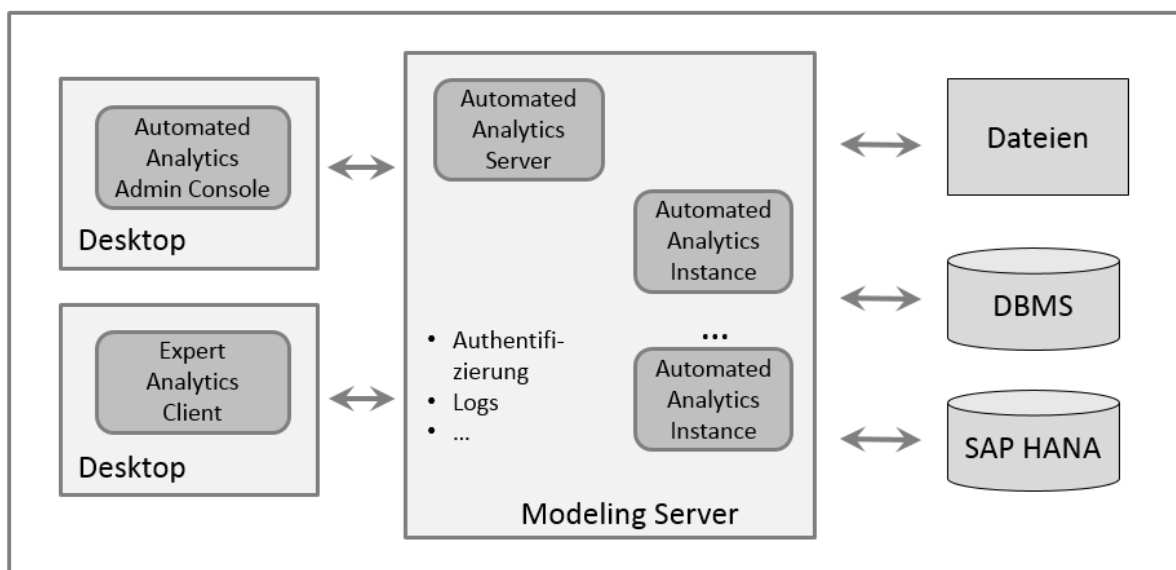


Abbildung 5: SAP Predictive Analytics – Enterprise-Edition

Die folgenden Erfahrungen wurden erhoben:

CRISP-DM	Bewertung
<p>Datenexploration</p>	<ul style="list-style-type: none"> - Klare Vorgabe in dieser Architektur ist die Nutzung des SAP PA Clients im Automated oder Expert Mode. Data Scientisten arbeiten heute jedoch vorherrschend mit Open Source Sprachen, insbesondere R und Python. Die Integration ist hier unzureichend (bzw. bei Python nicht unterstützt). Da sich Data Scientisten in ihrer Arbeitsweise nicht umstellen werden, betrachten Anwenderunternehmen die mangelnde Integration als einen Schwachpunkt - Die Arbeitsweise von Data Scientisten ist stark von einem interaktiven Umgang mit den Daten geprägt. Hierfür wäre der Expert Mode das richtige Werkzeug. Allerdings fehlen hier wesentliche Integrationsfunktionen mit HANA, wie z. B. <ul style="list-style-type: none"> - Ansprechbarkeit von Calculation Views, - automatische Typkonvertierungen zwischen Verarbeitungsknoten, - automatische Tabellenanlage, - automatisch sinnvoller Umgang mit Rechenlast, wenn die Prozesse die Kapazität überschreiten (z. B. Suspending von Jobs bei Engpässen, kein Abbruch, automatische Auslagerung von Prozessen auf kostengünstige Bereiche der HANA-Plattform, bei Anmeldung durch einen Data Scientisten)
<p>Datenvorbereitung</p>	<ul style="list-style-type: none"> - Anwender wollen hier den Data Manager und die Knoten im Expert Mode verwenden. Tatsächlich kommen dabei immer wieder sehr generische Fehlermeldungen z. B. zu fehlerhaften Datentypen (insbesondere Datentyp Decimal), so dass die Werkzeuge nicht nur nicht nutzbar, sondern auch nicht präsentierbar sind. Zwar sind zu konvertierende Datentypen z. B. im APL-Guide dokumentiert. Diese Vorgehensweise entspricht aber in keiner Weise der flexiblen und kreativen Arbeitsweise von Data Scientisten - Data Scientisten arbeiten in ihrer Datenvorbereitung mit häufigen Wertableitungen und Zwischenpersistierungen. Der Fokus liegt dabei auch nicht immer auf technischer Effizienz und performanter Programmierung, sondern auf der schnellen Erzeugung des inhaltlichen Ergebnisses und der flexiblen Kombination unterschiedlicher Datenquellen. Hierfür ist der Umgang mit HANA jedoch nicht ausgelegt (z. B. bei Datenvolumen-abhängiger Lizenz)

Modellierung	<ul style="list-style-type: none"> + Automated Analytics wird grundsätzlich als hilfreich empfunden - Allerdings fehlt die sinnvolle Verzahnung der Phase der Modellbildung mit derjenigen der Datenvorbereitung. Im Rahmen der Datenvorbereitung suchen Data Scientisten nach Eigenschaften in Daten, um weitere Prädiktoren zu finden (sog. Feature Engineering). Dies macht den Hauptteil der Arbeit der Data Scientisten aus. Des Weiteren präferieren die Data Scientisten im Hause den Funktionsreichtum verfügbarer Python Bibliotheken. APL und PAL werden hier nicht als ernsthafte Alternative gesehen
Bewertung	<ul style="list-style-type: none"> - Auch dieser Arbeitsschritt wird gerne mit etablierten Werkzeugen durchgeführt. Eine fehlende Integration z. B. mit R und Python ist hier ein Hindernis
Bereitstellung	<ul style="list-style-type: none"> - Die Überführung in den Regelbetrieb der IT, welcher SAP-basierend ist, ist die eigentliche Herausforderung. Zielsetzung ist hier das systematische Deployment, von Python- und R-Logiken der Data Scientisten in den technischen Betrieb innerhalb der SAP (z. B. Copy/Paste von Skript-Bestandteilen). Die Modelle sollen dann automatisiert einplanbar sein und hinsichtlich von Versionen und verwendeten Datenquellen in einem Repository verwaltbar

+ positive Bewertung
- negative Bewertung

Die folgenden Erfahrungen wurden erhoben:

CRISP-DM	Bewertung
Datenexploration	- Die Datenexploration wird durch diesen Ansatz nicht unterstützt. Diese muss de facto weiter in anderen Werkzeugen erfolgen, z. B. R-Studio
Datenvorbereitung	- Die Datenvorbereitung ist in der Data Science ein hochgradig interaktiver Arbeitsschritt, der eng mit der Modellierung verzahnt ist. Durch den technischen Integrationsansatz existiert nur die Möglichkeit, Datenvorbereitung auf Basis von Stichprobendaten in anderen Werkzeugen (z. B. R-Studio) zu betreiben. Verfolgt man diesen Ansatz, muss die Datenvorbereitung aber bei Inbetriebnahme in HANA recodiert werden. Alternative wäre, das Data Scientisten direkt auf HANA entwickeln. Dies ist aber von Data Scientisten im Fachbereich kaum zu erwarten
Modellierung	- Die Modellierung ist eng mit der Datenvorbereitung verbunden. Die Modellierung erfolgt hier in R, die Datenvorbereitung aber in HANA Skript. Die enge Verbindung der beiden Arbeitsschritte kann so nicht umgesetzt werden
Bewertung	- Der Modellvergleich wird typischerweise in R durchgeführt. So wie im Falle der Modellierung ist es aufgrund der Trennung der Werkzeugwelten erforderlich, HANA Studio und R zu verwenden und mit jedem Modellvergleich, R-Skripte in HANA zu integrieren. Dies erfordert vom Data Scientisten Kenntnisse beider Werkzeugwelten sowie die Bereitschaft, sich auf eine datenbankorientierte Philosophie im Umgang mit Daten einzulassen. Dies ist nicht realistisch
Bereitstellung	- Für die Implementierung sind stets IT-Eingriffe erforderlich. Dadurch ist das Verfahren für Fachbereiche nicht handhabbar. Wenn Veränderungen durchgeführt werden müssen, ist auf relativ niedriger Ebene individuell zu entwickeln und es ist entsprechend aufwändig. - Es existiert keine Unterstützung für automatisches Deployment, ein Versionsmanagement und keine Transportmöglichkeiten

+ positive Bewertung

- negative Bewertung

Anmerkung: Die R-Integration wurde insgesamt nicht sehr positiv bewertet. Dennoch wurde von Anwenderunternehmen auch das Problem einer fehlenden Python-Integration betont. Dem ist SAP inzwischen entgegengekommen. So lassen sich heute Predictive Analytics Algorithmen mit Python abfragen, um z. B. die wichtigsten Prädiktoren lt. SAP PA mit Python auszulesen.

5. AUSBLICK

Es hat sich gezeigt, dass der beim automatisierten Modus verwendete SAP-Algorithmus mit hoher Genauigkeit arbeitet, denn alle Ergebnisse lagen auf einem guten Leistungsniveau, verglichen mit den in die Tiefe gehenden Ansätzen. Hier liegt die Stärke des Tools.

Für die Zukunft hat SAP eine deutliche Fokussierung auf die SAP Analytics Cloud angekündigt. Dies betrifft auch SAP Predictive Analytics. So ist hier z. B. ein deutlich einheitlicheres Frontend zu erwarten, wie auch eine neue Oberfläche für den Data Scientisten mit dem Produktnamen „Predictive Composer“.

Unternehmen werden ihre riesigen Datenbestände nicht alle in der Cloud haben. Daten befinden sich auch zu sehr großen Teilen „außerhalb“ der SAP-Landschaften und SAP-Systemen.

Mit dem anstehenden Wechsel und der Nutzung der Unternehmen von S/4HANA, aber auch BW/4HANA muss eine tiefe Integration möglich sein.

Im Umfeld einer Cloud- und On-Premise-Strategie der Unternehmen muss das Predictive Analytics Tool hybride Systemlandschaften nutzen und sich in diese integrieren können.

6. ANHANG

SAP-Dokumentation

[SAP20160913]

SAP HANA R Integration Guide, SAP HANA Platform SPS 12 Document Version: 1.1
2016-09-13

[SAP20171026]

SAP BusinessObjects Predictive Analytics 3.1
2017-10-26

SAP Help

https://help.sap.com/viewer/p/SAP_PREDICTIVE_ANALYTICS

Bücher

https://www.sap-press.com/sap-predictive-analytics_4491/

https://www.rheinwerk-verlag.de/sap-predictive-analytics_4276/

Open SAP Kurse

<https://open.sap.com/courses/ds1>

<https://open.sap.com/courses/ds2>

Survival Guide Installation

<https://help.sap.com/viewer/c27db6dddfb04b2e8a4e44ad1a14ffa1/3.3/enUS/b8cf04dc266e4995a82090211d29fca3.html>

SAP Predictive Analytics: Quickly Preparing Data for Predictive Modeling

<https://www.youtube.com/watch?v=F-OCrM04D5I>

How does Automated Analytics do it? The magic behind creating predictive models automatically

<https://blogs.sap.com/2015/07/01/how-does-automated-analytics-do-it-the-magic-behind-creating-predictive-models-automatically/>

A Great Combination: Python and SAP Predictive Analytics

<https://blogs.sap.com/2017/12/21/a-great-combination-python-and-sap-predictive-analytics/>

Ferner soll die neue UI5-Maske lt. SAP deutlich intuitiver sein als der Desktop und auch eine neue integrierte Hilfe anbieten.

Neue Schulungsunterlagen sind lt. SAP aktuell in Bearbeitung.

Abbildungen

Abbildung 1

[SAP20171026]

Abbildung 2

https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining,

Abruf 26.3.2018

Abbildung 3

SAP BUSINESSOBJECTSPREDICTIVE ANALYTICS, Reimagine Predictive for The Digital Enterprise, Jan Fetzer, Product Management Advanced Analytics, 06.09.2016

Abbildung 4

[SAP20171026]

Abbildung 5

[SAP20171026]

Abbildung 6

[SAP20160913]

Weiterführende Informationen

www.r-project.org

en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

Embedding Predictive Analytics into SAP S/4HANA Cloud

<https://blogs.sap.com/2017/05/16/embedding-predictive-analytics-into-sap-s4hana/>

Enriching SAP BW with SAP Predictive Analytics – Introduction

<https://blogs.sap.com/2017/10/17/enriching-sap-bw-with-sap-predictive-analytics-introduction/>

Machine Learning Automation: Beyond Algorithms

<https://www.sap.com/documents/2016/02/c4d70a6c-617c-0010-82c7-eda71af511fa.html>

<https://blogs.sap.com/2017/12/21/a-great-combination-python-and-sap-predictive-analytics/>

7. IMPRESSUM

Wir weisen ausdrücklich darauf hin, dass das vorliegende Dokument nicht jeglichen Regelungsbedarf sämtlicher DSAG-Mitglieder in allen Geschäftsszenarien antizipieren und abdecken kann. Insofern müssen die angesprochenen Themen und Anregungen naturgemäß unvollständig bleiben. Die DSAG und die beteiligten Autoren können bezüglich der Vollständigkeit und Erfolgsgeeignetheit der Anregungen keine Verantwortung übernehmen.

Die vorliegende Publikation ist urheberrechtlich geschützt (Copyright).

Alle Rechte liegen, soweit nicht ausdrücklich anders gekennzeichnet, bei:

Deutschsprachige SAP® Anwendergruppe e.V.
Altrottstraße 34 a
69190 Walldorf | Deutschland
Telefon +49 6227 35809-58
Telefax +49 6227 35809-59
E-Mail info@dsag.de
www.dsag.de

Jedwede unerlaubte Verwendung ist nicht gestattet. Dies gilt insbesondere für die Vervielfältigung, Bearbeitung, Verbreitung, Übersetzung oder die Verwendung in elektronischen Systemen/digitalen Medien.

© Copyright 2018 DSAG e.V.