

Predictive Analytics und Big Data im Lichte moderner Umfragesysteme der IBM, oder was hat Predictive Analytics mit Karl Valentin zu tun?

von Dr. Christian Trippner
IBM, Senior Technical Sales Professional, Predictive Analytics

„Prognosen sind schwierig, vor allem wenn sie die Zukunft betreffen“

In diesem fast hundert Jahre alten Zitat vom Großmeister des hintergründigen Humors und Dada-Künstlers *Karl Valentin* steckt viel Wahrheit und es hat an Relevanz und Aktualität für die Gegenwart nichts verloren.

Prognosen und der Blick in die Zukunft beschäftigen Unternehmen und Organisationen bereits seit vielen Jahrhunderten, verschaffen sie doch in der Regel einen Wettbewerbsvorteil und versetzen Unternehmen in die Lage pro-aktiv zu handeln und Geschäftsprozesse präziser und erfolgreicher zu steuern. Fehlten in vergangenen Zeiten jedoch meist die Möglichkeiten und die passende Methodik, stehen dem Analysten und Data Scientisten heutzutage ein fast unüberschaubares Feld von Verfahren, Möglichkeiten und Werkzeugen zur Verfügung.

Die IBM Workbench SPSS Modeler bahnt einen Weg durch diesen Dschungel und stellt seit

Figure 3 The Forrester Wave™: Big Data Predictive Analytics Solutions, Q2 '15



über einem Jahrzehnt ein leistungsfähiges und optimal skalierbares Werkzeug dar, um diesen Aufgaben gerecht zu werden.

Wie gut die IBM mit ihrem Predictive Analytics Portfolio bezüglich Strategie, Vision und aktuellem Angebot aufgestellt ist zeigt sich in der nebenstehenden Grafik und dem Zitat.

Abbildung 1: Gartner Grafik der Unternehmen im Bereich Predictive Analytics

*"No matter how an organization wants to get started with predictive analytics, IBM has an option for them. The solution offers the most comprehensive set of capabilities to build models, conduct analysis and deploy predictive applications both on-premises and in the cloud."*¹

¹ The Forrester Wave™: Big Data, Predictive Analytics Solutions, Q2 2015

Wie erfolgreich die Tools und Programme der IBM SPSS Analyseplattform ist und welchen Mehrwert Anwender aus den Programmen ziehen, zeigt die ROI Studie von Nucleus Research. Hier wurden die Kunden und Anwender der SPSS Software nach abgeschlossenen Projekten zum Erfolg und den Erfahrungen mit SPSS Analyse- und Befragungssoftware befragt.

So sind ein überwiegender Teil der Projekte mit einer positiven Bilanz hinsichtlich Kosten, ROI (Return on Investment), Zeitplan und Budget, sowie der Projektlaufzeit beurteilt worden.

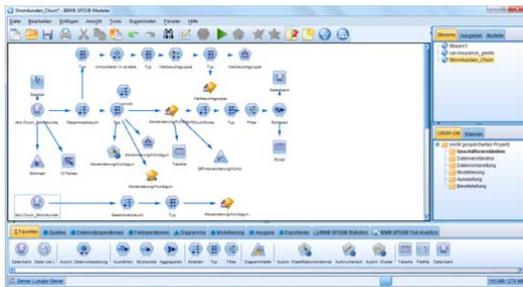
Nucleus Research ROI Report über SPSS

- 94% aller Projekte mit positivem ROI
- 10,7 Monate für die Erreichung des ROI (im Ø)
- 81% der Projekte im Zeitplan
- 75% der Projekte im oder unter Budget



Abbildung 2: ROI Studie Nucleus Research

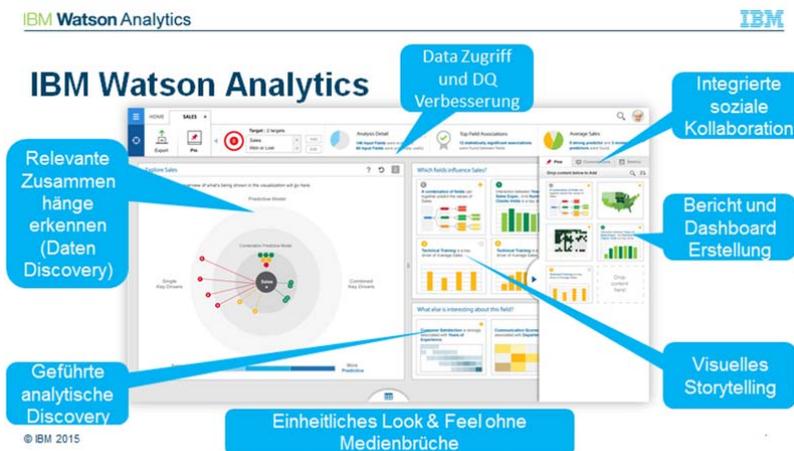
Der IBM SPSS Modeler ist ein modular aufgebautes Werkzeug, das neben den klassischen



Modellierungsalgorithmen (Klassifikation, Segmentierung und Assoziation) auch eine hochleistungsfähige Text Mining Workbench enthält und damit Unternehmen ein wertvolles Instrument in die Hand gibt, mit dem unternehmenskritische Entscheidungen, besser, schneller und kostengünstiger getroffen werden können.

Abbildung 3: Die IBM SPSS Modeler Workbench

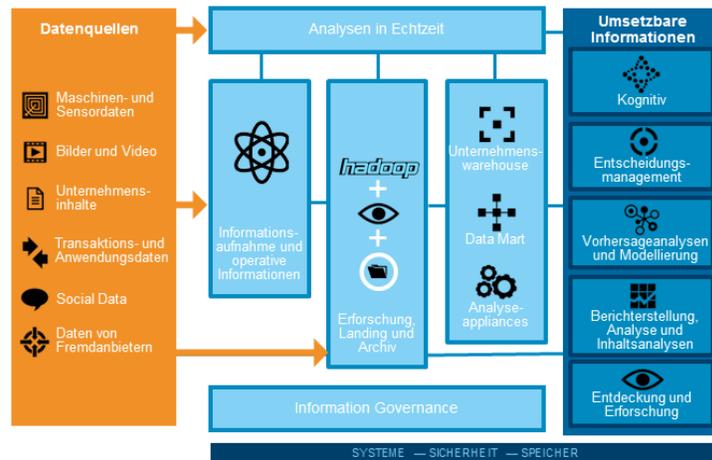
Mit Predictive Analytics und dessen Herzstück Data Mining können Unternehmensdaten gewinnbringend ausgewertet werden, in dem durch passende Algorithmen und bei Bedarf



dem Einsatz von Big Data Technologien (z. B. Hadoop und/oder Streaming Data), Muster, Strukturen und Trends in den Daten erkannt werden.

Abbildung 4: Benutzeroberfläche IBM Watson Analytics

Einen besonderen Akzent setzt in diesem Zusammenhang die intuitiv-intelligente Watson Analytics-Anwendung, die in der Lage ist, selbständig nach Mustern und Strukturen in Daten zu suchen und per Spracheingabe Analysen und Modellierungen durchführt („Zeige mir den Zusammenhang und die vorhandenen Muster zwischen Exportquote, Bruttosozialprodukt und Arbeitslosenquote für die Staaten in Europa mit mehr als 20 Millionen Einwohnern“).



Der Einsatz von Big Data Technologien ist im Kontext von Predictive Analytics besonders gewinnbringend, da hiermit sonst nicht oder nur teilweise nutzbare Massendaten, teilweise auch unstrukturiert (Texte, e-Mails, Kommentare, Blogs, ...) einer d Mehrwert stiftenden Nutzung zugeführt werden können.

Abbildung 5: Synoptischer Überblick - Big Data Analytics

Dieses Wissen kann dann in Unternehmensprozesse in „real-time“ eingebunden werden um am sogenannten Customer-Touch-Point bzw. im Produktionsumfeld, unmittelbar (moment of truth) wissensbasierte Prognosen abzugeben. Weiter dient Predictive Analytics zur periodischen bzw. saisonalen Steuerung von spezifischen Aktionen in den verschiedenen Fachabteilungen eines Unternehmens. Das können viele verschiedene Themen sein, z. B. Responsequotenoptimierung, Marketing- und Werbemaßnahmen, Kundenbindungsmaßnahmen und CRM Aktionen, Vorhersage von Abwanderung / Kündigung und Migration, Betrugserkennung, Qualitätssicherung, und so weiter und so fort, die Liste könnte fast endlos weitergeführt werden.

Predictive Analytics dient also vorrangig der Unterstützung und Präzisierung von unternehmerischen Entscheidungen.

Es stellt sich in diesem Kontext jedoch die Frage, ob in einem Unternehmen zu jeder Zeit und in jedem Fall, die für eine präzise Vorhersage und stabile Modellierung nötigen Daten überhaupt vorhanden sind.

Unsere Zeit ist geprägt durch den steten Wandel: Meinungen, Stimmungen, Vorlieben, Trends und Aktivitäten können sich extrem schnell ändern, angetrieben vor allem und auch durch das stetige Wachstum sozialer Netzwerke und sozialer Medien. Wer hier etwas verschläft, kann wirtschaftlich und technologisch schnell im Abseits stehen bzw. nur mehr durch teure Maßnahmen den Anschluß wieder finden.

„Unternehmen, die früh in Befragungssysteme investieren um Kundeninput zu erfassen, zu analysieren und dementsprechend zu agieren, werden einen klaren Vorteil am Markt haben“

Dieses Zitat stammt von keinem geringeren als Bill Gates² und zeigt wie wichtig das direkte und zeitnahe Erfassen von Daten sein kann.

Auch hier bietet die IBM eine umfassende und leistungsfähige Softwareplattform an, die den Anwender in die Lage versetzt über alle Befragungswege, die Zielgruppe genau dort abzuholen wo sie sich befindet und so Daten zu sammeln, die für eine erfolgreiche Unternehmung entscheidend sein können.



Ganz gleich ob Kunden befragt werden sollen, Lieferanten, Mitarbeiter oder eine andere Zielgruppe, SPSS Data Collection bietet Ihnen immer die passende Lösung. Dabei kann der Anwender frei zwischen den jeweiligen Befragungskanälen wählen: schriftlich, online, telefonisch oder mobil.

Abbildung 6: Umfragen mehrsprachig gestalten

Die Befragungsplattform ist modular aufgebaut und ermöglicht den parallelen Betrieb der verschiedenen Befragungswege, so daß Sie Ihre Zielgruppe immer treffend ansprechen können. Dabei wurde höchster Wert zum einen auf die Einfachheit der Bedienung gelegt, zum anderen können auch sehr anspruchsvolle Befragungsprojekte realisiert werden, da die Plattform vollständig in bereits existierende Unternehmenslösungen eingebunden werden kann. Die Data Collection Plattform ermöglicht den problemlosen Wechsel von einem Befragungskanal zu einem anderen bzw. ermöglicht die Erweiterung um einen weiteren Kanal (z. B. Papier und Online) und das ganz ohne Programmieraufwand.

Die nachfolgende Grafik gibt einen Überblick zu den in der IBM SPSS Data Collection verfügbaren Modulen und deren Einbindung in die weiteren Analysewerkzeuge von SPSS.

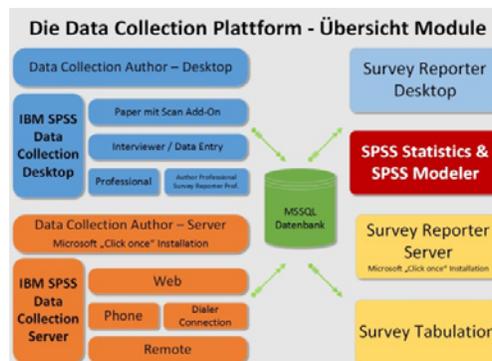


Abbildung 7: Übersicht zur IBM SPSS Data Collection Plattform

² Business @ the Speed of Thought by Bill Gates and Collins Hemingway, 1999

Ein Fragebogen wird mit dem IBM SPSS Author entworfen, der es erlaubt ohne jegliches Scripting oder andere Vorkenntnisse einen Fragebogen attraktiv zu erstellen und zu gestalten.

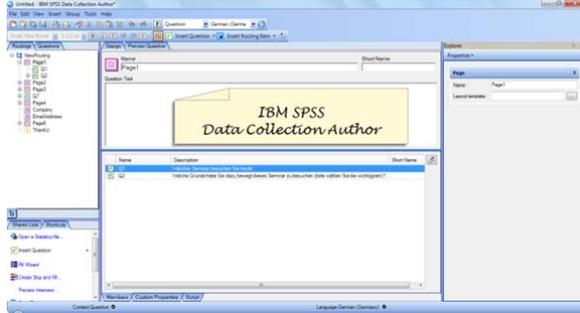


Abbildung 8: Der IBM SPSS Data Collection Author

Es wird für jedes Umfrageprojekt nur ein Fragebogen benötigt. In diesem sind dann alle möglichen Kanäle (schriftlich, online, telefonisch, ...), Ablaufversionen (Routings) und Sprachversionen enthalten. Im Author können bei der Fragebogenerstellung bereits wichtige, Einstellungen vorgenommen werden. So werden im Fragebogen bereits die fehlenden Werte, die Wertelabels und andere auswertungsrelevante Einstellungen festgelegt. Nach der Fertigstellung wird der Fragebogen dann im jeweiligen Modul verwendet.

Für technisch komplexere Umfragen steht dem Anwender auch eine leistungsstarke Umgebung für Scripting und Programmierung zur Verfügung.

Für schriftliche Befragungen auf Papier gibt es ein Modul, daß als Plug-In in MS Word® und bei Bedarf auch mit Anbindung an einen Scanner für ein automatisches Erfassen der Bögen, den klassischen Weg der PAPI-Befragung ermöglicht.



Q1 Welches Seminar besuchen Sie heute:

- Seminar 1
- Seminar 2
- Seminar 3

Q2 Welche Gründe habe Sie dazu bewegt dieses Seminar zu besuchen (bitte wählen Sie die wichtigsten)?

- Termin
- Schulungszentrum
- Dozenten
- Themen und Schwerpunkte

Abbildung 9: Papierfragebogen

Gerade Umfragen die personensensible Informationen erheben, bei denen ein Höchstmaß an Anonymität gefordert ist (z. B. Mitarbeiterbefragungen) werden nach wie vor gerne auf diesem Wege durchgeführt.

Ein weiteres Modul dient der lokalen Datenerfassung, z. B. auf einem oder mehreren Desktoprechner(n). Durch eine klar strukturierte Datenerfassungsmaske, in der auch Validierungen, Sprünge und Autoausfüllen möglich sind, können die Daten bequem eingegeben werden. Sind ganze Teams von Datenerfassern am Werk kann über die Definition von Batch-Jobs eine zeitliche bzw. hierarchische Staffelung der Eingabesessions erfolgen („multi-modal project mit batching und bundeling“). Des weiteren können in diesem Modul auch persönliche und lokale Umfragen (CAPI) offline auf einem Rechner durchgeführt werden, die Umfrageseiten mit einem bestimmten Layout verwenden sollen. Gerade wenn es um die Anreicherung Ihrer bestehenden Daten mit persönlichen

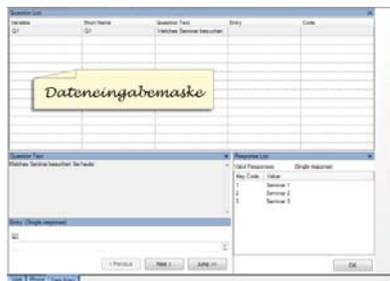


Abbildung 10: Data Entry Eingabemaske

Angaben geht oder bei erklärungsbedürftigen Umfragen, ist diese Art der Datenerhebung von Vorteil.

Sehr häufig werden Umfragen auch über das Inter- oder Intranet durchgeführt. Besonders attraktiv macht diese Befragungsform die Direktheit, die Möglichkeit mit dem Befragten interaktiv zu agieren, das schnelle Erreichen von Rücklauf und damit die Verfügbarkeit der Antworten. Die erhobenen Daten sind von Anfang an in einem digitalen Format. Durch technische Spielereien und eine entsprechende grafische



Abbildung 11: Administrationsoberfläche Web Interviews

Gestaltung, kann eine Umfrage sehr attraktiv und für jede Zielgruppe angepaßt erstellt werden. Besonders hervorzuheben ist hier auch der Umstand, daß der Anwender die mit IBM SPSS Data Collection erzeugten Umfragen auch auf eigenen Servern durchführen kann und somit immer die Hoheit über die gesammelten Daten besitzt.

Die Umfragen können dabei allgemein zugänglich gemacht werden oder mittels Sample Management nur einem bestimmten Benutzerkreis zur Verfügung gestellt werden.

Eng verwandt mit Onlinebefragungen sind Erhebungen über das Telefon. Auch hier bietet



Abbildung 12: CATI Eingabeoberfläche

die Data Collection Plattform eine äußerst leistungsfähige und professionelle Lösung. Neben einer klaren und gut strukturierten Benutzeroberfläche, die dem Agent als Arbeitsoberfläche dient, hat das System zahlreiche Möglichkeiten die im Telefonbetrieb wünschenswert sind: Terminvereinbarung, Wiedervorlage bei „besetzt“,

Einhaltung von zulässigen Anrufzeiten, Ausschlüsse bei der Terminvereinbarung, Interviewer-Queing, VOIP, festlegen von Interviewer-Skills und ein umfangreiches Reporting, sowie Auswertungsstatistiken sind nur die wichtigsten Punkte in diesem Zusammenhang.

Sowohl die Onlinebefragungen (CAWI) als auch die telefonischen Befragungen (CATI) können über eine praktische Administrationsoberfläche in einem HTML Browser verwaltet werden. Dabei sind neben der reinen Projektverwaltung mit Datenexport, Rücklaufkontrolle, Benutzerverwaltung und Sample Management, auch mandantengesteuerte Anwenderhierarchien und Rollenvergabe verfügbar.



Nicht immer steht für eine Befragung das notwendige Netz oder eine Internetverbindung zur Verfügung. Hier

Abbildung 13: Umfrage Mobil für Android

kann die mobile Befragungslösung der IBM SPSS Data Collection Plattform zum Einsatz kommen. Egal ob sie völlig offline, teilweise offline/online oder online sind, werden die Antworten der Befragung zuerst auf dem mobilen Gerät gesammelt und dann je nach Möglichkeit gebündelt synchronisiert und auf ein Datenhaltungssystem übertragen. Seit der neuesten Version 7.0.1 ist es auch möglich Umfragen über Smartphones mit Android OS durchzuführen.

Wenn es bei den gesammelten Daten dann um eine Analyse geht, können in bewährter

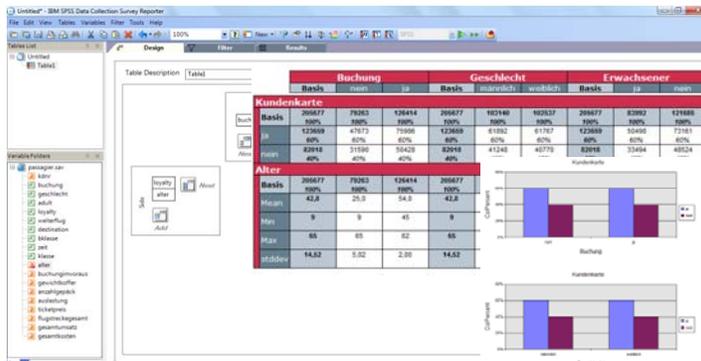


Abbildung 14: Der IBM SPSS Data Collection Survey Reporter - Tabellen und Grafiken

Form die klassischen SPSS Tools IBM SPSS Statistics oder der IBM SPSS Modeler genutzt werden. Aber auch aus der Plattform heraus gibt es Möglichkeiten Analysen, Tabellierungen und Reports zu erstellen. Die dazugehörigen Produkte sind der Data Collection Survey Reporter. Dieses Programm kann auf sehr einfache und selbsterklärende Weise genutzt werden,

um schöne deskriptive Tabellen zu erstellen, χ^2 und t-Tests durchzuführen und um anschließend einen automatischen Export in z. B. MS Power Point® oder MS Excel® anzustoßen.

Fazit:

Mit der IBM SPSS Data Collection Befragungsplattform ist der IBM ein ganz großer Wurf gelungen, der in seiner Komplexität weltweit wohl als einzigartig anzusehen ist, deckt die Plattform doch alle denkbaren Umfrageszenarien und Umfragekanäle ab. Dabei besticht die Lösung zum einen durch ihre Einfachheit in der Bedienung und der Schnelligkeit mit der Ergebnisse erzielt werden, zum anderen aber auch durch ihre Komplexität und Skalierbarkeit wenn es um technisch sehr anspruchsvolle Projekte geht.

Mit entsprechender Programmierung kann die Data Collection an praktische jede Anforderung und Einsatzumgebung angepaßt werden.

Die Auswertungs- und Tabellierungswerkzeuge der Plattform runden dabei das Gesamtbild ab und fügen sich perfekt und nahtlos in das gesamte SPSS Analyseportfolio ein.

Davon zeugen auch viele nationale und internationale Projekte und Kunden, die die Plattform in den unterschiedlichsten Branchen und Anwendungsgebieten nutzen.

Die Plattform wird eingesetzt bei

- Bürgerbefragungen im öffentlichen Sektor,
- Fahrgastbefragungen an Flughäfen und im ÖPNV,
- Umfragen für und an Konsumenten,
- Produktstudien,
- wissenschaftliche Erhebungen,
- Justiz- und Vollzugsbehörden zur Qualitätssicherung,
- öffentlicher Rundfunk und Fernsehen,
- Logistikunternehmen,
- Produktentwicklung und viele andere Bereiche mehr.

Weitere Informationen und detaillierte Produktbeschreibungen zu der IBM SPSS Umfrageplattform Data Collection erhalten sie unter www.ibm.com/de/spss