

Der Weg zu Business Intelligence

Comelio

Der Weg zu Business Intelligence

Business Intelligence ist ein Schlagwort, das durchaus für eine Reihe von verschiedenen Verfahren stehen kann und das von einer starken Methodenvielfalt gekennzeichnet ist, die unterschiedliche Ansprüche erfüllen und auch verschiedene Zielsetzungen haben.

■ BI im Unternehmen

■ Zum Begriff Business Intelligence

Als deutschen Begriff kann man Geschäftsanalytik verwenden, wobei das Wort Intelligenz hier nicht im eigentlichen deutschen Sinne von Klugheit oder Geistesstärke zu sehen ist, sondern vielmehr in den Methoden, wie Daten gesammelt und untersucht werden. Die übergeordnete Zielsetzung dieser Methoden ist dabei durch den zusätzlichen Begriff Business vorgegeben, wobei die Methodik wie Statistik oder Data Mining

durchaus nicht nur betriebswirtschaftlichen Fokus besitzt, sondern in Natur- und Sozialwissenschaften genau so zum Einsatz kommen. Diese Methodik wird allerdings so eingesetzt, dass gesteckte Unternehmensziele mit notwendigen operativen und strategischen Entscheidungen erreicht werden, die durch die Ableitung von auf Basis dieser Methodik gewonnenen Erkenntnissen basieren.

■ Erfolgreicher und effektiver Einsatz von BI

Ernstgemeintes Business Intelligence ist durchaus mehr als nur die Einführung von Standardwerkzeugen beliebiger Hersteller wie Microsoft, Oracle oder SAP. Immer noch ist gerade die fachlich versierte Unterstützung und die methodenorientierte Herangehensweise entscheidend für den tief greifenden Erfolg von BI-Projekten und fest installierten BI-Prozessen.

Oftmals beschränkt man sich damit, ein strukturiertes Berichtswesen einzuführen, das nicht selten hauptsächlich die Daten aus dem Leistungserstellungsprozess aufgreift und sie dann in traditionellen Berichten oder auch multidimensional aufzubereiten.

Erfolgreiches Business Intelligence erfordert im Regelfall eine eigenständige Datenbasis, die aufgrund ihrer Größe typischerweise als Data Warehouse (DW oder DWH) bezeichnet wird, wobei grundsätzlich die Methoden, die BI begründen, auch auf Basis von kleinen Datenmengen und auch ohne den Einsatz einer solchen Datenbasis erfolgreich zum Einsatz kommen können und in einer solchen Konstellation oftmals auch in kleinen Einzel-

projekten sehr gut verwertbare Ergebnisse liefern können. Die Zusammenführung von Daten in einem gemeinsamen Data Warehouse ist normalerweise dann erforderlich, wenn verschiedene Datenquellen die Einzeldaten bereit stellen und diese Daten zunächst in einem eigenständigen ETL-Prozess (Extraktion-Transformation-Laden) zusammengeführt werden müssen.

Die Bedeutung von Business Intelligence ließ sich insbesondere im Jahre 2007 gut erkennen, als solche Hersteller von betriebswirtschaftlicher Software und Datenbanken wie Oracle und SAP bislang eigenständige Unternehmen mit eigenen Produkten in diesem Bereich übernommen haben und nun diese Produkte in ihre bestehende Anwendungsarchitektur integrieren oder selbstständig weiter entwickeln. Aber auch die Entwicklung von eigenen Lösungen im Bereich Expertensysteme sowie eine Reihe von Open-Source-Produkten für Data Mining-Auswertungen zeichnet ein lebhaftes Marktbild mit ständigen Veränderungen und permanenter Neuentwicklung.

■ BI und Datenbeschaffung durch Befragungen

Das hier thematisierte und im Vordergrund stehende Thema externe Datenerhebung durch Befragungen und Beobachtungen bietet dabei im Schritt der Datenerfassung nach der planenden Modellbildung und dem allgemeinen Design des Untersuchungsvorhabens eine logische Vorarbeit für Data Mining und Wissensentdeckung. Wie schon zuvor erläutert, können die Daten, die im Rahmen von klassischen statistischen Auswertungen oder auch mit den Methoden des Data Mining untersucht werden, entweder aus schon vorhandenen Datenquellen stammen und müssen ggf. gereinigt oder mit anderen Daten angereichert werden, oder sie werden durch eine Befragung und damit einen eigenen Schritt der Datenerhebung ermittelt.

In diesem Sinne kann man sich vorstellen, wie bspw. Fragestellungen, die teilweise auch auf Basis der schon vorhandenen Daten beantwortet werden können, um weitere ergänzende Fragen intensiviert werden. So ist es dann möglich, nicht nur Tendenzen in der Verkaufsentwicklung, die Unterschiede von Produktionsqualität im Laufe der Zeit oder ähnliche Themengebiete aus dem Bereich der Leistungserstellung zu betrachten, sondern für diese Themengebiete zusätzlich auch sozialwissenschaftliche oder psychologische Aspekte mit einzubeziehen.

Insbesondere der menschliche Faktor kann hier immer wieder als gängiges Beispiel für weitergehende Untersuchungen und ergänzende Modellbildung herangezogen werden. Durch eine Befragung bei Kunden, Lieferanten und Mitarbeitern lassen sich Motive und

Gründe für Verhaltensänderungen oder über einen längeren Zeitraum dauerhafte oder sich in der Zukunft sogar verändernde Verhaltensmuster ableiten und erklären. Mit diesen Erkenntnissen ist es dann wiederum möglich, andere zu beobachtende Phänomene wie die Arbeitssicherheit, Kunden- und Mitarbeiterzufriedenheit, Kundenbindung oder erwartbare Phänomene wie zukünftiges Bestellverhalten, Vertragskündigungsquoten sowie Akzeptanz und Reaktion von Veränderungen im Produkt- und Servicebereich vorherzusagen und sie zu unterstützen oder ihnen entgegen zu steuern.

Der Unterschied zwischen Business Intelligence, das ausschließlich auf Daten beruht, die aus dem Leistungserstellungsprozess und hier insbesondere den tatsächlich eingetroffenen Bestellungen, die Absatz- und Marktstruktur, wie sie sich in schon erfolgten Bestellungen, Stornierungen oder auch Reklamationen widerspiegelt, oder auch allgemein die Produktivität, die nur den tatsächlichen Output, nicht aber beeinflussende Faktoren wie Mitarbeitermotivation und Organisationsstruktur berücksichtigt, und einer Vorgehensweise wie Business Intelligence, das zusätzlich auch sozialwissenschaftliche Aspekte berücksichtigt, kann in qualitativer Hinsicht gar nicht groß genug eingeschätzt werden. Plakativ könnte man festhalten, dass hier sozusagen trockene Zahlen, die im Wesentlichen auch nur aus der Vergangenheit stammen und allerhöchstens einen gegenwärtigen Zustand abbilden, mit Hilfe einer Befragung auch so dimensioniert werden

können, dass sie vor einem sozialen Hintergrund überhaupt erst eine Tiefe erhalten, mit der eine vollständige Erklärung von bspw. vergangenem, gegenwärtigem und zukünftigem Kundenverhalten greifbar wird. Es lassen sich hier ganz andere Modelle zur Erklärung und – was gerade für ein Unternehmen einen besonderen Reiz darstellt – auch für die Prognose wie schließlich auch der Steuerung von Verhaltensweisen entwickeln. Typische Vorteile der Datenauswertung sind hier im Wesentlichen wiederum dieselben, wie sie auch schon bei den reinen Daten der Leistungserstellung aufgezeigt werden könnten, aber – so die hier dargestellte Idee – nicht wirklich in vollem Umfang eingelöst werden können: bessere Ausrichtung von Werbemaßnahmen, Anpassung von Produkt- und Servicequalität an nachgewiesene Erwartungen, Verbesserung von betrieblichen sozialen Gegebenheiten zur verbesserten Erreichung der übergeordneten Unternehmensziele.

Als Kerngedanke lässt sich festhalten, dass sich durch die Hinzunahme von Methoden der empirischen Sozialforschung, zu denen insbesondere die Befragung/Beobachtung gehören, in bestimmten Bereichen eine soziale Dimension aufspannen lässt, mit denen einfachen Daten aus dem Leistungserstellungsprozess erst das notwendige Leben eingehaucht werden kann, um sie nicht nur darzustellen, sondern ihr Zustandekommen und ihre Ausprägungen auch erklären zu können.

Selbstverständlich – auch das soll nicht unerwähnt bleiben – findet dieser Ansatz seine Grenzen dort, wo überhaupt gar keine sozia-

le Dimension für eine Modellentwicklung notwendig ist. Dies ist bspw. der Fall bei eindeutig im technischen Bereich liegenden Untersuchungsbereichen. Ein Untersuchungsgegenstand, der im Wesentlichen mit Hilfe von technischen (physikalischen oder chemischen bspw.) Erklärungsansätzen gedeutet werden kann, lässt sich nicht durch eine Befragung unter den beteiligten Mitarbeitern oder den Kunden des Endprodukts so um eine soziale Dimension erweitern, dass die Produktionsqualität positiv beeinflusst wird. Wenn nun einmal die Rezeptur einer Substanz oder die Fertigungstechnik suboptimal sind, dann wird dies nur mit einer entsprechenden Hinwendung zu den technischen Gegebenheiten deutlich und kann auch nur in dieser Dimension verbessert werden. Jedoch – und dies muss auch in diesem Fall sofort wieder erwähnt werden – ist zu hinterfragen, ob möglicherweise in der sozialen Dimension hätte aufgedeckt werden können, dass durch eine veränderte Organisationsstruktur, die Mitarbeiteranregungen und Verbesserungsvorschläge begünstigt oder sogar darauf aufbauend ein Anreizsystem installierbar machen könnte, schon viel früher kostenintensive Rückrufaktionen oder allgemein eine hohe Ausschussquote hätte vermieden werden können. Gerade dieser Gedanke kann wiederum verdeutlichen, dass in einem sozialen Umfeld, wie es nun einmal ein Unternehmen nach innen und nach außen darstellt, psychologische und soziale Faktoren in mannigfaltiger Weise miteinander so verquickt sind, dass sich an der Oberfläche auch Auswirkungen in scheinbar anderen Dimensionen ausbilden können.

■ Ausgangspunkt: Traditionelle statistische Auswertungen

In diesem Abschnitt geht es um die Darstellung von BI und die Vorteile, die einem Unternehmen entstehen, wenn es zusätzliche Daten in Form von Befragungen erhebt, welche schon bestehende Datenmengen noch um eine entscheidende weitere Dimension ergänzen. Auch der Begriff Data Mining ist schon gefallen, wie verschiedene Andeutungen zur Modellbildung auf der Grundlage einer systematischen Datenauswertung. Dabei ist Data Mining durchaus nicht die Karte, die als erstes gezogen wird. Vielmehr handelt es sich dabei um fortgeschrittene Techniken, die Muster und Zusammenhänge erkennen können, welche selbst wiederum traditionelle statistischen Techniken einsetzen, miteinander kombinieren und weiterentwickeln. Ein strukturiertes Berichtswesen ist aus einem BI-Kreislauf nicht wegzudenken und stellt in den meisten Situationen eine sichere Grundlage dar. Der einzige entscheidende Marketing-Nachteil, den diese Techniken aufweisen, liegt insbesondere darin begründet, dass kein griffiger und plakativer englischer Name dafür existiert.

Man unterscheidet zwei grundsätzliche Strömungen bei statistischen Methoden, mit denen man auch sehr gut in dieser kurzen übersichtlichen Darstellung operieren kann. Die deskriptive (beschreibende) Statistik bietet Techniken an, mit denen umfangreiches Datenmaterial beschrieben und in vor allen Dingen zusammenfassender und verdichtender Weise analysiert werden kann. Sie lässt sich auch als Vorstufe von der induktiven (schließenden) Statistik begreifen, die

auf Basis der möglicherweise schon getätigten Untersuchungen weitere Techniken einsetzt, um Schlussfolgerungen und Regeln über Zusammenhänge ableitbar machen.

In beiden Fällen und so auch bei der Datenerhebung per Befragung und bei der Verarbeitung mithilfe von Data Mining-Methoden führt man Messungen an Untersuchungseinheiten durch. Man beobachtet dabei im Rahmen von Messungen bestimmte Merkmalsausprägungen oder Merkmalswerte. Die Merkmalswerte lassen sich dann in die beiden klassischen Gruppen qualitativer und quantitativer Merkmale aufspalten, wobei die quantitativen sich durch ihre Größe und die qualitativen sich durch ihre Art unterscheiden. Gängige Beispiele für solche quantitativen Größen sind das Alter, die Körpergröße oder das finanzielle Einkommen, während gängige Beispiele für qualitative Größen im Geschlecht, dem Beruf oder Farben sind. Die Messung erfolgt an einer Skala, die verschiedene Skalenniveaus haben kann. Neben einer Reihe von unterschiedlichen Skalenniveaus lässt sich die metrische Skala sehr schön von der Ordinalskala unterscheiden. Werte auf einer metrischen Skala bieten die Eigenschaften, dass sie in eine Reihenfolge gebracht werden können und interpretierbare und damit mathematisch verarbeitbare Abstände untereinander aufweisen. Eine Ordinalskala dagegen erlaubt die Anordnung der Werte.

Vor diesem Hintergrund sollen nun kurz die beiden Hauptrichtungen der statistischen Verarbeitung skizziert werden.

■ Deskriptive Statistik

Die *deskriptive Statistik* (beschreibende Statistik) stellt Techniken bereit, das Datenmaterial übersichtlich darzustellen und insbesondere Tabellen und Graphiken als Präsentationsmittel zu verwenden. Für die grobe Kennzeichnung von Datenmengen existiert eine Reihe von Kenngrößen.

Zu diesen charakteristischen Werten gehören die Lage- und die Streuungsparameter oder auch die Konzentration von Werten. Diese sind weithin bekannt und haben dadurch eine hohe Akzeptanz, um als Ergebnisse für eine Untersuchung auch in einer Analyse oder einem Bericht zu erscheinen.

Als Beispiele für Lagemaße kommen das arithmetische Mittel/der Mittel-/Durchschnittswert, der Median/Zentralwert und der Modalwert (häufigster Wert) in Betracht. Als Beispiele für Streuungs-/Dispersionsmaße, mit denen man die Abweichung vom

Zentrum der Datenmenge beschreiben kann, lassen sich die Spannweite (Wertebereich einer Beobachtungsreihe), die Varianz (mittlerer quadrierter Abstand zum Mittelwert), die Standardabweichung (mittlerer Abstand zum Mittelwert) und der Variationskoeffizient (Verhältnis von Standardabweichung und Mittelwert).

Auch die Kennzahlensysteme/Gliederungs-/Verhältniszahlen gehören zur deskriptiven Statistik und erfreuen sich gerade im BI-Bereich besonderer Beliebtheit: Beziehungszahlen (Quotient zweier sich sachlich ergänzender Maßzahlen), Gliederungszahlen (Bezug zwischen einer untergeordneten zu einer übergeordneten Größe) und Indexzahlen (zwei Maßzahlen gleicher Art werden zueinander in Bezug gesetzt) gehören zu dieser Gruppe.

■ Induktive Statistik

Die *induktive Statistik* (schließende Statistik oder Inferenzstatistik) stellt Methoden zur Verfügung, mit denen man Schlussfolgerungen aus dem gewonnenen Datenmaterial ziehen kann. Im Wesentlichen handelt es sich dabei um solche Schlussfolgerungen, die

- Parameterpunktschätzungen, bei denen Parameter (Kennzahlen der Verteilung) der Grundgesamtheit als reelle Zahl aus den Daten der Stichprobe zu schätzen sind.
- Parameterbereichsschätzungen, bei denen ein Parameter der Grundgesamtheit auf

Die Grundlage für die verschiedenen auch über die drei genannten Punkte hinausge-

auf der Basis von Stichproben-Daten Erkenntnisse über die Grundgesamtheit liefern sollen.

Man unterscheidet hier drei wesentliche Bereiche, für die entsprechende Methoden zur Verfügung stehen:

- Basis der Stichprobendaten in einem Intervall mit einer vorgegebenen Wahrscheinlichkeit geschätzt werden soll.
- Testverfahren, bei denen zwischen zwei sich gegenseitig ausschließenden Hypothesen entschieden werden soll.

henden Methoden der induktiven Statistik ist die Wahrscheinlichkeitstheorie.

■ Relationale Datenbanken und der Weg zu Data-Mining

Man kann sich die Entwicklung von Datenbanken als eine fast zwangsläufige Entwicklung hin zu Data Mining-Techniken vorstellen. Vereinfacht könnte man die nachfolgenden Phasen benennen, wobei ihre Auswahl und Zusammenstellung insbesondere die Perspektive der Auswertung unterstützen.

■ Strukturierte Speicherung

Am Anfang stand überhaupt die Möglichkeit, Daten elektronisch und in einer Form zu sammeln, die es möglich machte, zu einem späteren bzw. auch zu einem sehr viel späteren Zeitpunkt wieder auf diese Daten mit unterschiedlichen Fragestellungen zuzugreifen

zu können. Hier stand also die reine Speichermöglichkeit und die Datenstrukturierung in unterschiedlichen Ausprägungen (hierarchisch, netzwerkartig und schließlich relational) sehr im Vordergrund.

■ Echtzeit-Zugriff

Technische Verbesserungen und insbesondere die fortgeschrittene Entwicklung von relationalen/tabellenorientierten Speichermodellen und ihre relativ schnelle Ausbreitung und breite Unterstützung auf Werkzeugebene bot dann die Gelegenheit, auch große Datenmengen für eine Vielzahl von konkurrierenden Benutzerzugriffen mit Transakti-

onsunterstützung anzubieten. In diese Phase fällt dann auch der Begriff bzw. das Konzept von OLTP (OnLine Transaction Processing), d.h. die Entwicklung und Bereitstellung von bspw. unternehmensweit genutzten Datenbanken mit entsprechenden Software-Anwendungen als Benutzerschnittstelle für die direkte Bearbeitung von Daten.

■ Strukturierte Berichtserstellung

Neben Lese- und Schreibzugriffen auf die Daten wurden Werkzeuge zur Produktreife oder als individuelle Lösungen entwickelt, um auf aggregierte Werte der Datenbank zuzugreifen und sich so Berichtssysteme mit zusammenfassenden und vergleichenden Informationen aufzubauen. Diese konnten und können auch heute entweder statische Berichte produzieren, die innerhalb eines von Berichtsperiode zu Berichtsperiode

gleich bleibenden Rasters tabellarische oder grafisch aufbereitete Daten-Zusammenfassung bis hin zu sehr vereinfachenden Kennzahlen präsentieren, oder fortgeschrittenen bzw. mit besonderen Berechtigungen ausgestatteten Benutzern die Möglichkeit bieten, selbstständige einfache Untersuchungen unter Einsatz von Summierungen, einfachen statistischen Auswertungen und natürlich Filter durchzuführen.

■ Multidimensionale Analyse

Durch wiederum fortschreitende Entwicklung im Bereich der Methodik und der technischen Infrastruktur wurde das Konzept von OLAP (OnLine Analytical Processing) entwickelt. Es bietet den Benutzern wiederum die Gelegenheit, aggregierte Zusammenstellungen der Daten einzusehen, wobei aber hier eine multidimensionale Aufbereitung der Daten eigenommen wird. Sie erlaubt es, die Daten aus vorab definierten Winkeln und entlang von Dimensionen und Hierarchien zu

betrachten. Wie schon bei der strukturierten Berichtserstellung auch kann man Benutzergruppen definieren, die standardisierte Auswertungen mit vorgefilterten und vorab ausgewählten Dimensionen und Hierarchie-Ebenen präsentiert bekommen, und solchen, die mit entsprechender Werkzeug-Unterstützung in die Lage versetzt werden, eigenständig Zusammenstellungen und Kombinationen von Dimensionen und interessierenden Maßzahlen aufzubauen.

■ Muster- und Wissensentdeckung

Bisweilen als Teil der multidimensionalen Analyse angesehen, aber durch Methodik und Standardwerkzeuge auch als eigenständiger Entwicklungsschritt anzusehen, erlaubt die Entdeckung von Mustern und neuem Wissen in Daten die Chance, nicht nur schon bekannte Informationen zu filtern und thematisch zu gliedern und multidimensional

aufzubauen, sondern Zusammenhänge und Abhängigkeiten zwischen Daten abzuleiten und sie für die nachträgliche Erklärung oder die Prognose von Wertausprägungen zu verwenden. Dies wird oft auch als KDD (Knowledge Discovery from Data) bezeichnet.

■ Übersicht: Von Relationalen Datenbanken zu Data Mining

