

**Technische Universität Dortmund
Fakultät Maschinenbau
Fachgebiet IT in Produktion und Logistik
Prof. Dr. Ing. Markus Rabe**

Bachelorarbeit

**Zuordnung von Data Mining-Methoden zu problemspezifischen
Fragestellungen von Supply Chain Management-Aufgaben**

**Ebru Gürez (141886)
Wirtschaftsingenieurwesen
Beckstr. 128, 46238 Bottrop
ebru.guerez@tu-dortmund.de**

**Erstprüfer: Univ.-Prof. Dr.-Ing. Markus Rabe
Zweitprüferin: Dipl.-Inf. Anne Antonia Scheidler**

Ausgegeben am 19.03.2015

Eingereicht am 10.06.2015

Inhaltsverzeichnis

| | |
|--|-----|
| Abbildungsverzeichnis..... | III |
| 1 Einleitung..... | 1 |
| 2 Planung und Steuerung von SCM..... | 3 |
| 2.1 Begriffliche Grundlagen des SCM..... | 3 |
| 2.2 Ziele und Aufgaben des SCM..... | 8 |
| 2.3 Produktionsplanung und -steuerung im SCM..... | 10 |
| 2.4 Fragestellungen im SCM und ihre Kategorisierung..... | 12 |
| 3 Knowledge Discovery in Databases | 15 |
| 3.1 Definition, Einordnung und Abgrenzung des KDD..... | 15 |
| 3.2 Beispielmodelle des KDD..... | 16 |
| 3.3 Data Mining als Phase im KDD-Prozess | 19 |
| 3.3.1 Generelle Beschreibung des Data Mining..... | 19 |
| 3.3.2 Gängige Methoden des Data Mining..... | 20 |
| 4 Einsatz von Data Mining Methoden im SCM | 28 |
| 4.1 Untersuchung der ausgewählten SCM-Fragestellungen und Data Mining- Methoden..... | 28 |
| 4.2 Systematische Gegenüberstellungen der eingesetzten Data Mining-Verfahren im SCM | 35 |
| 5 Exemplarische Fallbeispiele | 41 |
| 5.1 Fallbeispiel 1: Clusterverfahren im Beschwerdemanagement..... | 41 |
| 5.2 Fallbeispiel 2: Teilebedarfsprognose in der Automobil-industrie | 43 |
| 6 Zusammenfassung und Ausblick | 45 |
| Literaturverzeichnis | 47 |
| Anhang..... | 51 |

Abbildungsverzeichnis

| | |
|---|----|
| Abb. 1: Entwicklungsstufen des Supply Chain Managements | 4 |
| Abb. 2: Prozesselemente der Supply Chain | 5 |
| Abb. 3: Ebenen des SCOR-Modells | 7 |
| Abb. 4: Aufgaben des Supply Chain Managements | 9 |
| Abb. 5: KDD Stufenmodell | 17 |
| Abb. 6: Hierarchisches Clusterverfahren | 22 |
| Abb. 7: Partitionierendes Clusterverfahren | 22 |
| Abb. 8: Neuron | 24 |
| Abb. 9: Neuronales Netz mit variabler Anzahl verdeckter Schichten | 24 |
| Abb. 10: Entscheidungsbaum | 26 |
| Abb. 11: Cluster der Rücklieferungsgründe | 42 |
| Abb. 12: Neuronales Netz für den Teilebedarf | 44 |

1 Einleitung

Aufgrund einer zunehmenden Globalisierung und der damit einhergehenden Erhöhung der Wettbewerbsfähigkeit sehen sich immer mehr Unternehmen gezwungen, durch eine kooperative Zusammenarbeit ihre betrieblichen Prozesse effizienter und wirtschaftlicher zu gestalten. Diese kooperative Vernetzung von Unternehmen führt zu einem Wertschöpfungsnetzwerk, welches erheblich komplexer in seiner Gestaltung ist als ein allein agierendes Unternehmen. Hieraus resultiert eine wesentliche Aufgabe für das Supply Chain Management (SCM): bestehende Aktivitäten und Geschäftsprozesse zwischen den vernetzten Unternehmen zu planen, zu steuern und zu überwachen (Becker et al., 2008). Für die Erfüllung dieser Aufgabe ist ein gezielter Einsatz von Informationstechnologie (IT) erforderlich. In diesem Sinne kann man beobachten, dass die Verbreitung digitaler Werkzeuge zur Planung, Steuerung und Überwachung im SCM permanent zunimmt (Deuse et al., 2013). Allerdings fallen durch die wachsende Integration der IT in die Wertschöpfungskette immer mehr an Daten an. Diese zunehmende Datenbereicherung hat den Vorteil, dass detailliertere Beobachtungsdaten über die verschiedensten Bereiche entlang der Wertschöpfungskette, wie z.B. über Beschaffung, Produktion und Absatz, vorliegen. Durch die Nutzung von modernen Informationstechnologien wird zwar eine strukturierte digitale Speicherung dieser Datenbestände ermöglicht, jedoch liegt der Fokus selten in ihrer effizienten Analyse hinsichtlich der Entdeckung von nützlichen Informationen und neuem Wissen (Deuse et al., 2013). Für ein wettbewerbsfähiges und erfolgreiches SCM haben sich im Rahmen der Informations- und Wissensgewinnung verschiedenste Auswertungsmethoden, sogenannte Data Mining-Methoden, durchgesetzt. Ziel dieser Methoden ist es, durch eine effiziente Nutzung der angefallenen Daten nützliche Informationen zu gewinnen und neues Wissen über die relevante Umwelt zu erhalten (Wiedmann et al., 2001). Diese beiden Forschungsbereiche SCM und Data Mining werden bisher nur selten intensiv zusammen betrachtet. Es gibt wenige Untersuchungen, die den einzelnen Bereichen einer SCM mögliche Methoden des Data Mining zuordnen. (vgl. Gluchowski et al., 2008; Becker, 2011; Deuse et al., 2013)

Im Rahmen dieser Arbeit stellt der oben beschriebene Sachverhalt den Ausgangspunkt dar: Es soll untersucht werden, inwiefern aus großen Datenmengen, die in Wertschöpfungsnetzwerken erzeugt werden, die Fragestellungen des SCMs anhand geeigneter Data Mining-Methoden beantwortet und verarbeitet werden können. Das Ziel hierbei ist, ausgewählte Data Mining-Methoden und SCM-Aufgaben systematisch gegenüber zu stellen und miteinander zu verknüpfen, um so die Produktionsplanung und -steuerung effizienter und effektiver zu gestalten.

Hierzu wird zunächst ein besonderes Augenmerk auf die Planung und Steuerung im SCM gelegt, wobei in einem ersten Schritt die Grundlagen, Ziele und Aufgaben aufgearbeitet werden. Basierend auf den aktuellen Entwicklungstendenzen im SCM werden zudem anhand gängiger Literatur relevante Fragestellungen erfasst und kategorisiert. Des Weiteren erfolgt eine Einordnung und Abgrenzung des Data Mining im Rahmen des Knowledge Discovery in Databases-Prozess, wobei im Anschluss mögliche Methoden vorgestellt werden. Die oben beschriebenen Grundlagen werden schließlich durch eine systematische Gegenüberstellung ergänzt. Es erfolgt eine Zuordnung und Bewertung von gängigen Data Mining-Methoden zu relevanten Fragestellungen des SCM, sodass darauffolgend durch Fallbeispiele Möglichkeiten zur praktischen Umsetzung aufgezeigt werden.

2 Planung und Steuerung von SCM

Die unterschiedlichen Planungs- und Steuerungsmöglichkeiten im Supply Chain Management (SCM) sind in dieser Arbeit von hoher Relevanz. Aus diesem Grund ist es notwendig, zunächst die Grundlagen des SCM darzulegen sowie die Ziele und Aufgaben zu erörtern. Weiterhin werden die Planungs- und Steuerungsprozesse aufgezeigt, sodass im Anschluss die während dieser Prozesse aufkommenden Leitfragen spezifiziert und kategorisiert werden können.

2.1 Begriffliche Grundlagen des SCM

Im Laufe der Jahre hat das SCM verschiedene Entwicklungsstufen durchschritten und damit einhergehend sind vielfältige Definitionen zustande gekommen, die weitestgehend vom Grad ihrer Motivation und ihrem Interesse abhängen (Bolstorff et al., 2007). Es ist kaum möglich, eine allgemeingültige Definition des SCMs in der breiten Literatur zu erkennen (vgl. Harrington, 1995; Towill, 1996; Fisher 1997). Allerdings kann man laut Porter (2006, 2010) im Konsens annehmen, dass in den Definitionen „[...] der Gedanke einer Integration von Unternehmungsaktivitäten aufgegriffen wird.“ (Werner, 2013) Diese Unternehmungsaktivitäten beziehen sich zum einen auf Prozessaktivitäten, die sich innerhalb des Unternehmens vollziehen, und zum anderen werden sie als ein integrierter Teil eines Wertschöpfungsnetzwerks betrachtet. Sie beinhalten Aktivitäten von Versorgung bis Entsorgung, begleitend von Material-, Geld- und Informationsflüssen. (Werner, 2013)

Die durchschrittenen Entwicklungsstufen teilen sich nach Baumgarten (2004) in vier verschiedene Phasen auf:

- (1) Integration der Funktionen interner Supply Chains
- (2) Informationsaustausch zwischen Kunden, Lieferanten und Dienstleistern
- (3) Kollaboratives Management kompletter Netzwerke und
- (4) Synchronisation interner wie externer Supply Chains.

Im Zusammenhang mit der Integration der Funktionen interner Supply Chains sollen die unterschiedlichen Funktionsbereiche im Unternehmen, wie beispielsweise Einkauf, Vertrieb oder Produktion, miteinander verknüpft werden, sodass die aufgebauten Prozessketten die interne Kooperation verbessern können. In der zweiten Stufe Informationsaustausch zwischen Kunden, Lieferanten und Dienstleistern intensiviert sich die unternehmensübergreifende Kommunikation der Supply Chain Partner, wobei dieser Prozess durch den Einsatz moderner Informationstechnologien realisiert wird.

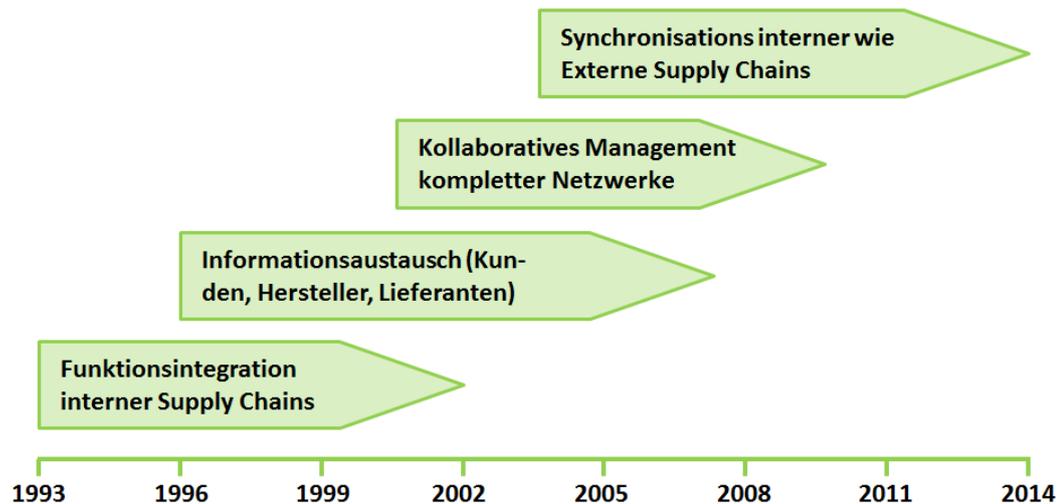


Abb. 1: Entwicklungsstufen des Supply Chain Managements (Baumgarten, 2004)

Um mögliche Synergiepotentiale auszuschöpfen, werden zu diesem Zweck Wertschöpfungsallianzen gebildet. In Anlehnung an die Ergebnisse der zweiten Phase, „[...] werden im *kollaborativen Management kompletter Netzwerke* [die] Kommunikation und Verteilung von Informationen in Echtzeit angestrebt.“ (Hilpert, 2014) Des Weiteren sollen Simultanplanungskonzepte eingesetzt werden, die für eine durchgängige Gestaltung des Informationsflusses zuständig sind. Nach Baumgarten (2004) spiegelt die vierte und letzte Entwicklungsstufe des SCMs *Synchronisation und Reduzierung interner wie externer Supply Chains* einen praktisch noch nicht vollständig erreichten Zustand der heutigen Supply Chains wieder, die sich als hoch komplex, kompliziert und intransparent erweisen. Kennzeichnend hierfür ist der Zuwachs der am Supply Chain teilnehmenden Unternehmen und damit einhergehend die erschwerte Planung und Steuerung der Wertschöpfungskette. In diesem Sinne wird betont, dass die Nutzung von Informationstechnologie immer mehr an Bedeutung gewinnt. (Baumgarten, 2004)

In der heutigen Zeit versuchen Unternehmen die bestehende Komplexität, Kompliziertheit und Intransparenz im SCM mit verschiedenen Ansätzen zu steuern. Das von dem Supply Chain Council (SCC, 2011) entwickelte und bis heute vorangetriebene Supply Chain Operation Reference (SCOR)-Modell ist ein standardisierter und branchenübergreifender Ansatz, welcher die in Wertschöpfungsketten entstehenden Prozesse und Aktivitäten systematisch beschreibt und definiert. Demzufolge stellt das SCOR-Modell ein Beschreibungsmodell dar und dient somit nicht für eine ideale Ausgestaltung der Supply Chain. (Eßig et al., 2013) Weiterhin vergleicht das SCOR-Modell definierte Managementprozesse mit Benchmarkingdaten, Best Practices und Softwarefunktionalitäten. (Bolstorff et al., 2007)

Als Geschäftsprozess-Referenzmodell erstreckt sich der SCOR Ansatz über die gesamte Wertschöpfungskette, „[...] das heißt alle Material-, Waren- und Informationsströme

vom Lieferanten des Lieferanten zur Fertigung über die Auslieferung bis hin zum Kunden des Kunden“ (Beckmann, 2012). Das SCOR-Modell setzt sich aus vier hierarchisch strukturierten Ebenen zusammen (Bolstorff et al., 2007; Eßig et al., 2013):

- **Ebene 1 (Höchste Ebene):** In dieser Ebene wird festgelegt, welchen Umfang und Inhalt die Supply Chain aufweisen soll. Hierbei unterscheidet man grundsätzlich zwischen den folgenden fünf Prozesselementen: Planen, Beschaffen, Herstellen, Liefern und Rückliefern. Diese Elemente charakterisieren die Lieferkette und durch ihre Verknüpfung werden die entstehenden Beziehungen zwischen Kunden und Lieferanten definiert.

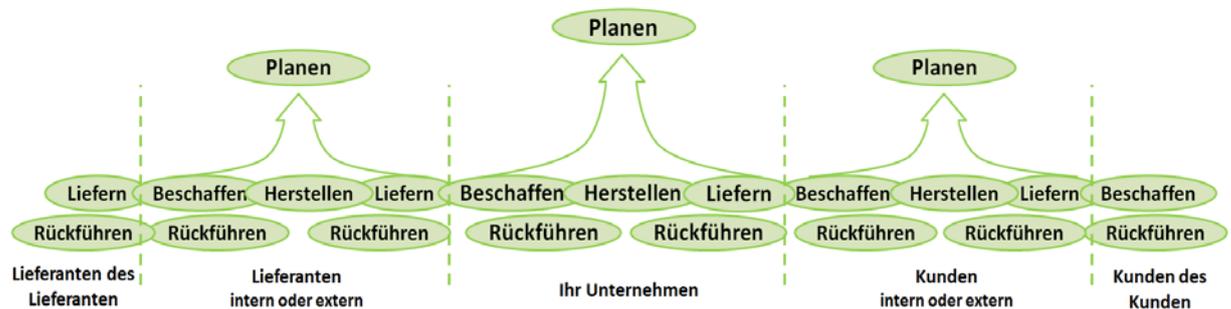


Abb. 2: Prozesselemente der Supply Chain (Bolstorff et al., 2007)

Das Prozesselement *Planen* beinhaltet Aktivitäten wie den Zugriff auf angebotsseitige Ressourcen, die Aggregation und Priorisierung von nachfrageseitigem Bedarf, die Lagerbestandsplanung zur Befriedigung von Absatz-, Produktions- und Materialbedarfe sowie die grobe Kapazitätsplanung für alle Produkte und Absatzkanäle.

Das *Beschaffen* umfasst den Bezug von Waren, Wareneingang, Eingangskontrolle, Lagerung und Zahlungsanweisung für Rohmaterialien und des Weiteren noch für bezogene Fertigprodukte.

Der Prozess des *Herstellens* umfasst die Materialbedarfsanforderungen und den Materialeingang, die Produktion und das Prüfen von gefertigten Produkten, die Zwischenlagerung oder die Freigabe zur Auslieferung von Fertigprodukten.

Im Prozesselement *Liefern* sind Tätigkeiten wie das Abwickeln von Aufträgen, das Erstellen von Angeboten, die Produktkonfiguration, das Anlegen und Pflegen von Kunden-, Artikel- sowie Preisstammdaten von Wichtigkeit. Weiterhin umfasst dieses Prozesselement das Verwalten von Forderungen, Krediten, Rechnungsstellung und Zahlungseingängen, das Ausführen von Lagerprozessen einschließlich Artikelkonfiguration, die Warenentnahme und -

verpackung, das Erstellen von kundenspezifischen Verpackungen und Auszeichnungen, die Auftragszusammenführung, den Warenversand, die Steuerung und Überwachung des Transportprozesses sowie den Ein- und Ausfuhr.

Das letzte Prozesselement des *Rücklieferns* beinhaltet die Abwicklung von rückgesendeten Artikeln mit Defekten unter Gewährleistung und bei Überbelieferung, inklusive die Freigabe, die Terminierung, die Inspektion, die Beförderung, die Verwaltung der Gewährleistungsbestimmungen, den Empfang und die Prüfung defekter Produkte sowie die Disposition und Ersatzlieferung.

- **Ebene 2 (Konfigurationsebene):** In dieser Ebene werden die oben beschriebenen fünf Prozesselemente in Teilprozesse zerlegt, sodass es möglich ist, Prozesse innerhalb der Supply Chain zu konfigurieren. Hierzu werden die Teilprozesse drei verschiedenen Prozesstypen zugeordnet: (1) Planungsprozesse, (2) Ausführungsprozesse und (3) Ermöglichungsprozesse. Prozesse, die Angebot und Nachfrage ausgleichen und diese Aktivität innerhalb eines bestimmten Planungszeitraumes durchführen, werden den Planungsprozessen zugeordnet. In Ausführungsprozessen sind Prozesse enthalten, die durch einen ausgelösten Planungsprozess zu einer Transformation führen und somit den Status des Produktes verändern. Prozesse, die zur Informationsaufbereitung oder Datenpflege dienen und somit eine Voraussetzung für die Planungs- und Ausführungsprozesse sind, werden den Ermöglichungsprozessen zugeordnet. Durch diese Verknüpfung und Interaktion von Prozessen und Prozesstypen wird ermöglicht, eine Prozessmatrix mit allen Kombinationsmöglichkeiten zu erstellen, sodass die Unternehmung die zutreffende Prozesskette und ihre geeignete Konfiguration erhält.
- **Ebene 3 (Gestaltungsebene):** In der Gestaltungsebene wird das SCOR-Modell weiter konkretisiert um eine branchenspezifische Referenzmodell-Konfiguration ermöglichen zu können. Diesbezüglich werden die Prozesselemente $P X_i$ (mit $i = 1, \dots, n$) definiert und ihre Input-Output-Relationen ermittelt. Weiterhin werden auch Leistungskennzahlen und Benchmarks festgelegt, sodass Unterschiede und Defizite zu den Best Practices erkennbar sind. Hierzu werden

die benötigten Systemfähigkeiten und erforderliche Softwareanwendungen spezifiziert.

- **Ebene 4 (Implementierungsebene):** Aufgrund von unternehmensinternen und –spezifischen Verfeinerungen der Prozesselemente, wird die Implementierungsebene nicht als eine direkte Ebene SCOR-Modells angesehen. Trotzdem ist die Implementierungsebene notwendig, da der SCOR-Ansatz sonst nicht vollendet durchgeführt wird.

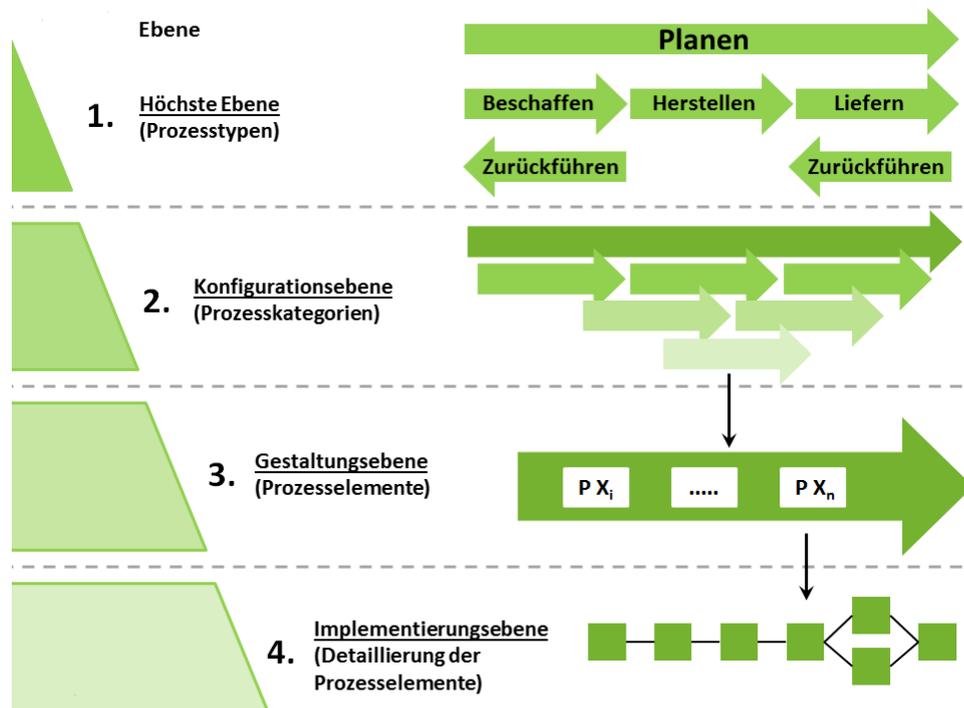


Abb. 3: Ebenen des SCOR-Modells (Bolstorff et al., 2007)

Zusammenfassend ist festzuhalten, dass das SCOR-Modell ein verbreiteter Ansatz zur Analyse und Verbesserung der Lieferkette ist. Durch eine einheitliche Sprache zwischen den unterschiedlichen unternehmensinternen Funktionsbereichen und den Unternehmenspartnern im Supply Chain Netzwerk wird über die gesamte Lieferkette hinweg eine optimierte Gestaltung der Kunden-Lieferanten-Beziehungen ermöglicht. Das SCOR-Modell sollte jedoch nicht als ein alleinstehender Ansatz zur Realisierung einer erfolgreichen Lieferkette gesehen werden. Insbesondere in Kombination mit anderen Führungsmodellen, wie z.B. dem Veränderungsmanagement oder der Geschäftsprozess-Optimierung, kann der SCOR-Ansatz zu einem langfristigen Unternehmenserfolg führen. (Hieronimus, 2005; Bolstorff et al., 2007)

2.2 Ziele und Aufgaben des SCM

Die übergeordnete Zielvorstellung im SCM ist verbunden mit der Befriedigung aller Kundenbedürfnisse, während die Kosten im Unternehmensnetzwerk im Hinblick auf Bestände, Ressourcen und Prozesse gleichzeitig optimiert werden sollen. Dadurch wird eine sichere und sich fortlaufend weiterentwickelnde Wettbewerbsfähigkeit des Netzwerkes angestrebt. Dieses übergeordnete Ziel teilt sich in mehrere verschiedene Teilziele auf:

- Verbesserung der Kundenzufriedenheit
- Erhöhung der Liefertreue und Lieferbereitschaft
- Verkürzung der Prozess- und Durchlaufzeiten
- Verringerung der Bestände entlang der Wertschöpfungskette
- Flexibilisierung von Produktions- und Logistikprozessen
- Verbesserung der Planungsqualität
- Erhöhung der Produktivität

Mit Hilfe dieser Optimierung des Gesamtsystems wird zum einen eine verbesserte Kundenorientierung angestrebt und zum anderen werden die Wettbewerbspotentiale des Wertschöpfungsnetzwerks ausgenutzt. (Hellingrath, 2008)

Werner (2013) betont in diesem Kontext, dass sich das Erfolgspotential des SCM aus folgenden vier Wettbewerbsfaktoren ergibt, die ähnlich stark auszurichten sind: Kosten, Zeit, Qualität und Flexibilität. Der Faktor Kosten bezieht alle kostenverursachenden Parameter wie Bestände, Frachten, Investitionen oder Abschreibungen auf logistische Assets ein. Weiterhin wird die hohe Relevanz von Zeit insbesondere bei der Reduzierung von Durchlaufzeiten und Time-to-Market, d.h. in dem Zeitraum bis neue Produkte marktreif sind, ersichtlich. Der Faktor Qualität wird in den Fällen wie Ausschuss oder Nacharbeit gemessen und fordert die Befriedigung der Kundenansprüche. Der Faktor Flexibilität hingegen drückt die Wandlungs- und Anpassungsfähigkeit des Unternehmens aus. (Arndt, 2010; Werner, 2013)

Um die gewünschten Ziele erreichen zu können, muss das SCM sowohl innerhalb eines Unternehmens als auch unternehmensübergreifend gewisse Aufgaben durchführen. Zu den wesentlichen Aufgaben im SCM gehören die Versorgung, die Entsorgung und die Wiederverwertung der integrierten Unternehmungsaktivitäten (Werner, 2013). Für eine adäquate Zielerreichung differenzieren Kuhn und Hellingrath (2002) zwischen drei Hauptebenen im SCM: Gestaltungsebene (Supply Chain Design), Planungsebene (Supply Chain Planning) und Ausführungsebene (Supply Chain Execution). Wichtig hierbei ist zu erwähnen, dass sich dieses SCM-Aufgabenmodell aus dem SCOR-Modell ableiten lässt (Werner, 2013).

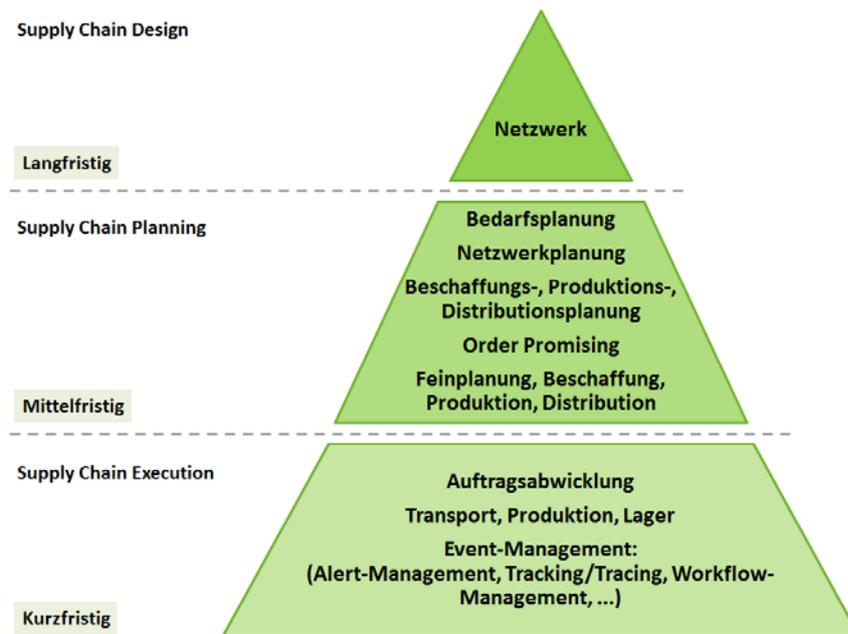


Abb. 4: Aufgaben des Supply Chain Managements (Hellingrath, 2008)

Auf der strategischen Gestaltungsebene erfolgen langfristig anhaltende Planungen, die sich mit Strukturierungsmöglichkeiten der Supply Chain befassen. Hierzu gehören z.B. Aspekte wie Standortplanung von Produktionsstätten, Umfang und räumliche Anordnung von Lagern sowie die Wahl geeigneter Kooperationspartner. Während diesen Planungen sollte darauf geachtet werden, dass das gesamte Logistiknetzwerk im Hinblick auf die entstehenden Kosten, so effektiv wie möglich gestaltet und ausgelegt wird. (Kuhn/Hellingrath, 2002)

Die zweite Ebene umfasst den Bereich der taktischen und operativen Planungsebene. Es werden Entscheidungen bezüglich der Bedarfe, Bestände, Kapazitäten oder Kapazitätszuordnungen der Akteure getroffen. Das Ziel dieser Ebene ist unter Berücksichtigung von kapazitäts- und terminbedingten Abhängigkeiten eine passende Programmplanung zu erstellen, die sich über einen mittel- bis langfristigen Zeithorizont erstreckt. Im Rahmen der Bedarfsplanung liegt der Schwerpunkt auf verschiedenen Prognoseverfahren, wobei hier zwischen kurz-, mittel- und langfristigen Bedarfsermittlungen unterschieden wird. Während bei einem kurzfristigen Bedarf die vorhandenen Bestellungen ausgewertet werden, wird bei einem mittelfristigen Bedarf der Marktbedarf auf Produktebene ermittelt. Beim langfristigen Bedarf hingegen erfolgt eine Prognose über den Absatz der Produktfamilien für einen längeren Zeitraum. Die Netzwerkplanung hat die Aufgabe unternehmensübergreifend die einzelnen Netzwerkpartner im Wertschöpfungsnetz zu koordinieren. Durch diese übergreifende Koordinierung gelingt es den Netzwerkpartnern Bedarfe sowie Kapazitäts- und Materialressourcen aufeinander abzustimmen. Die Beschaffungsplanung orientiert sich an der Zielvorstellung, die richtigen Güter am richtigen Ort zum richtigen Zeitpunkt zu

beschaffen, wobei die Lagerbestände zu minimieren sind. Hauptaufgabe der Produktionsplanung ist es, für jeden einzelnen Standort der Supply Chain einen Produktionsplan zu erstellen und diese zu optimieren. Hierbei wird das Ziel verfolgt, Faktoren wie die Termintreue und Lieferbereitschaft zu maximieren, Kapazitätsauslastungen zu optimieren und Bestandskosten zu minimieren. Im Rahmen der Distributionsplanung liegt die Aufgabe darin, die Lagerbestände optimal zu planen und einen sicheren Warenstrom zum Kunden zu gewährleisten. Beim Order Promising handelt es sich um Verfügbarkeits- und Machbarkeitsanalysen von eingehenden Kundenanfragen, wobei zwischen folgenden Ansätzen unterschieden wird: (1) Available-to-Promise (ATP), (2) Capable-to-Promise (CTP) und (3) Configure-to-Promise (CoTP).

Im Sinne des ATP-Ansatzes wird überprüft, ob das gewünschte Gut im Lagerbestand verfügbar ist. Der CTP-Ansatz setzt dahingegen voraus, dass das Unternehmen die Fähigkeit besitzt, das gewünschte Produkt produzieren zu können. Demzufolge wird hier überprüft, ob die Kapazitäten und Materialien für eine Produktion ausreichend vorhanden sind. Beim CoTP-Ansatz wird das gewünschte Produkt nach dem Kundenwunsch konfiguriert und auch hier erfolgt eine Prüfung der notwendigen Kapazitäten und Materialien. Des Weiteren erfolgt in dieser Ebene eine Feinjustierung der Beschaffungs-, Produktions- und Distributionsprozesse. (Kuhn/Hellingrath, 2002)

In der dritten und letzten Ebene liegt der Fokus auf der exekutiven Betriebsebene. Gegenstand dieser Ebene sind kurzfristige Entscheidungen über Kundenaufträge hinsichtlich der Auftragsplanung und ihrer logistischen Umsetzung. Hierbei hat die Auftragsabwicklung die Aufgabe die Bearbeitung von Kundenaufträgen zu steuern und zu überwachen. Sie umfasst damit auch die Lager-, Produktions- und Transportabwicklung der Aufträge. Ein wichtiger Bestandteil der exekutiven Ebene ist das Supply Chain Event Management. Es wird eine permanente Überwachung aller Supply Chain Aktivitäten beabsichtigt, sodass möglichst zeitnah alle Störungen, Bedarfe oder Engpässe identifiziert und für alle betroffenen Netzwerkpartner zugänglich gemacht werden können. (Kuhn/Hellingrath, 2002)

2.3 Produktionsplanung und -steuerung im SCM

Aufgrund der sinkenden Fertigungstiefe der Unternehmen und der damit einhergehend wachsenden Anzahl an Lieferanten haben sich die üblichen innerbetrieblichen Schwerpunkte der Produktionsplanung und -steuerung im Rahmen des SCM verschoben. Eine effektive unternehmensinterne Produktionsplanung ist nicht brauchbar, wenn die Lieferung der bestellten Güter nicht termingerecht erfolgt. Daher

müssen zwischenbetriebliche Zusammenhänge erarbeitet werden, sodass anstelle der eigenen die gesamte Lieferkette optimiert werden kann. (Kurbel, 2011)

Mittels Einsatz moderner Informationstechnologien ist über die gesamte Wertschöpfungskette eine funktionsübergreifende sowie simultane Planung und Steuerung möglich (Corsten/Gössinger, 2008). Dabei handelt es sich um sogenannte Advanced Planning and Scheduling (APS)-Systeme, die den optimalen Planungsprozess innerhalb der Supply Chain unterstützen (Van Eck, 2003). Dieses System basiert auf den Enterprise Resource Planning (ERP)-Systemen, die verschiedene Vorteile aufweisen, wie z.B. „[...] eine integrierte Datenbasis mit integrierten Zugriffsmöglichkeiten, Vermeidung von mehrfachem Datenerfassungsaufwand, Vermeidung von Fehlern bei der Transformation von Daten, kürzeren Zugriffszeiten und besseren Auswertungsmöglichkeiten der integrierten Daten“ (Lawrenz et al., 2000). Im Sinne einer optimalen Planung stellen die im ERP-System hinterlegten Datenbestände wie z.B. Stamm-, Planungs-, Wiederbeschaffungs- oder Kapazitätsdaten die Basis für ein funktionierendes APS-System dar. Mit diesen Daten ist das APS-System in der Lage sowohl eine Kapazitäts- und Bedarfsplanung durchzuführen als auch simultan komplexe logistische Strukturen zu visualisieren. (Grünauer et al., 2000; Meyr et al., 2008) Auf diese Weise wird ermöglicht, innerhalb des ganzen Wertschöpfungsnetzwerks einen synchron laufenden und aufeinander abgestimmten Planungsprozess zu realisieren. Diese Synchronisation der Planungsprozesse hat zur Folge, dass die Durchlaufzeit verkürzt wird, der Lagerbestand reduziert wird, Lieferzeiten eingehalten und Fertigungspläne realisierbar gestaltet werden. (Kuhn/Hellingrath, 2002)

Betge (2006) unterscheidet zwischen den folgenden sechs wesentlichen Merkmalen von APS-Systemen:

- **Kundenorientierung:** Produktion nach dem Pull-Prinzip
- **Realitätsnähe:** Verfügbarkeit von Beständen und Kapazitäten in Echtzeit; Vermeidung von Engpässen
- **Simultaneität:** Simultan ablaufender Planungsprozess
- **Geschwindigkeit & Flexibilität:** Hohe Verarbeitungsgeschwindigkeit der Planungsabläufe; Ausgeprägte Anpassungsfähigkeit auf Abweichungen
- **Integration:** Aufnahme von Lieferanten- und Kundeninformationen
- **Simulation:** Durchspielen von wenn-dann-Szenarien

Allerdings ist es wichtig, weiterhin eine hohe Datenqualität sicherzustellen und Mitarbeiter intensiv für den Umgang mit der Software zu schulen. Die alleinige Implementierung des APS-Systems ist nicht ausreichend für einen reibungslosen Ablauf der Produktionsplanungen. (Werner, 2013)

2.4 Fragestellungen im SCM und ihre Kategorisierung

Im Folgenden Abschnitt werden basierend auf verschiedene Literaturquellen spezifische Fragestellungen des SCMs abgeleitet. Hierbei dient das in der Unternehmenspraxis weit verbreitete und an verschiedene Supply Chains anpassungsfähige SCOR-Modell (vgl. hierzu Kapitel 2.1) als Strukturierungsmöglichkeit, da sonst verschiedene Fragen strukturlos aus unterschiedlichen Granularitätslevel resultieren. Durch die fünf Prozesselemente *Planen*, *Beschaffen*, *Herstellen*, *Liefern* und *Rückliefern* ist es möglich sämtliche Aktivitäten und Fragen, die im Rahmen des Wertschöpfungsprozesses vorkommen, abzubilden und auch zu kategorisieren.

Das Prozesselement Planen deckt nach Bolstorff et al. (2007) verschiedene Planungskomponenten wie Bedarfs-, Absatz-, Produktions- und Kapazitätsplanung ab. Weiterhin wird ebenfalls die Infrastruktur, insbesondere die Einrichtung von Distributionszentren, der Planung fest zugeschrieben. Diesbezüglich können folgende Fragestellungen herangezogen werden:

- Wie hoch werden zukünftige Kundenbedarfe sein?
- Wo entsteht der Bedarf? Bei welchen Kunden bzw. in welchen Regionen liegt dieser Bedarf vor?
- Welche Distributionszentren sollen wo eingerichtet werden und welche Kunden beliefern?

Durch die geringe Fertigungstiefe im Unternehmen und der damit verbundenen wachsende Fremdbeschaffung von Produktteilen, erreicht das Prozesselement des Beschaffens eine immer größere Bedeutung. In diesem Sinne sind Kriterien wie die Liefertreue, Lieferqualität und Lieferantenbewertung sowie die Menge zu beschaffender Produkte von hoher Relevanz. (Kurbel, 2011) Anhand dieser Kriterien können folgende Fragen entwickelt werden:

- Bei welchem Lieferanten soll bestellt werden?
- Welche Produkte sollen in welchen Mengen beschafft werden, um eine optimale Lagerverwaltung und einen optimalen Materialfluss zu gewährleisten?

Für den Sektor Herstellen sind laut Fandel et al. (2011) die Faktoren Produkt und Maschine besonders im Hinblick auf Qualität, Maschinenausfälle und Herstellungszeiten von hoher Bedeutung, sodass sich folgende Fragestellungen ableiten lassen:

- Wie hoch ist die Qualität der produzierten Teile?

- Wie oft müssen Wartungen an Produktionsanlagen durchgeführt um zukünftige Maschinenausfälle zu vermeiden?
- Wie lassen sich die Rüstzeiten optimieren?

Basierend auf Kiener et al. (2012) werden im Prozesselement Liefern Maßnahmen durchgeführt, die zur Befriedigung der Kundennachfrage dienen. Hierzu gehören z.B. die Kommissionierung, die Verpackung sowie der Versand der Ware. Diesbezüglich können folgende Fragestellungen herangezogen werden:

- Welches Transportmittel oder welcher Transportweg ist wann am effizientesten?
- Wie kann man die Anzahl der verspäteten Aufträge minimieren und eine fristgemäße Lieferung sicherstellen?

Hinsichtlich des Rückliefern bzw. Zurückführens stehen laut Bolstorff et al. (2007) unter anderem folgende Kriterien im Mittelpunkt: die Abwicklung von rückgesendeten Artikeln mit Defekten unter Gewährleistung, Empfang und Prüfung defekter Produkte, Disposition sowie Ersatzlieferung. Anhand dieser Kriterien lassen sich folgende Fragestellungen ableiten:

- Worauf lässt sich der Defekt bzw. die Beschwerde am rückgelieferten Produkt zurückführen?

Nachfolgend werden tabellarisch alle den Prozesselementen zugeordneten Fragestellungen zusammengefasst:

Tabelle 1: Relevante Fragestellungen des SCM

| Prozesselement | Nr. | Fragestellung |
|-------------------|-----|--|
| Planen | 1 | Wie hoch werden zukünftige Kundenbedarfe sein? |
| | 2 | Wo entsteht der Bedarf? Bei welchen Kunden bzw. in welchen Regionen liegt dieser Bedarf vor? |
| | 3 | Welche Distributionszentren sollen wo eingerichtet werden und welche Kunden beliefern? |
| Beschaffen | 4 | Bei welchem Lieferanten soll bestellt werden? |
| | 5 | Welche Produkte sollen in welchen Mengen beschafft werden, um eine optimale Lagerverwaltung und Materialflüsse zu gewährleisten? |

| | | |
|--------------------|-----------|---|
| Herstellen | 6 | Wie hoch ist die Qualität der produzierten Teile? |
| | 7 | Wie oft müssen Wartungen an Produktionsanlagen durchgeführt um zukünftige Maschinenausfälle zu vermeiden? |
| | 8 | Wie lassen sich die Rüstzeiten optimieren? |
| Liefern | 9 | Welches Transportmittel oder welcher Transportweg ist wann am effizientesten? |
| | 10 | Wie kann man die Anzahl der verspäteten Aufträge minimieren und eine fristgemäße Lieferung sicherstellen? |
| Rückliefern | 11 | Worauf lässt sich der Defekt bzw. die Beschwerde am rückgelieferten Produkt zurückführen? |

3 Knowledge Discovery in Databases

Die Daten- und Informationsgewinnung entlang der Wertschöpfungskette ist eine wesentliche Grundlage für ein wettbewerbsfähiges und erfolgreiches SCM. Aufgrund der wachsenden Integration der IT in die Wertschöpfungskette, entstehen große Datenbestände, die durch eine gezielte Datenanalyse zur Wissensentdeckung beitragen können (Gabriel et al., 2009). Ein zu diesem Zweck entwickelter Ansatz stellt der Knowledge Discovery in Databases (KDD)-Prozess dar. Das Ziel hierbei ist es, durch eine effiziente Nutzung und Verarbeitung der angefallenen Datenbestände neues Wissen über die relevante Umwelt zu gewinnen. (Fayyad et al., 1996) Im Rahmen dieses Kapitels werden die Grundlagen des KDD-Ansatzes erläutert und ein mögliches Beispielmodell aufgezeigt. Dabei werden basierend auf der Themenstellung der vorliegenden Arbeit das Data Mining als eigene Phase des KDD-Modells aufgegriffen und gängige Methoden vorgestellt.

3.1 Definition, Einordnung und Abgrenzung des KDD

Der anwachsende Datenbestand in einem Unternehmen und die daraus resultierende Möglichkeit, neue und implizit auffindbare Wissensentdeckungen durchzuführen, haben dazu beigetragen, dass der Prozess der Datenanalyse zunehmend an Bedeutung gewinnt. Einen beispielhaften Ansatz der Datenanalyse stellt der KDD-Prozess dar. Hierbei wird KDD definiert als „[...] non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [...]“ (Fayyad et al., 1996). Basierend auf einem nichttrivialen Prozess werden Datenmuster ermittelt, die einen realen Zusammenhang zwischen den beobachteten Merkmalen modellieren. Der Begriff nichttrivial (*non-trivial*) drückt hierbei aus, dass es sich nicht um einfache statistische Auswertungen von Daten oder reine Datenbankabfragen handelt sondern komplexe Such- oder Schlussfolgerungsalgorithmen angewendet werden. Des Weiteren müssen die ermittelten Datenmuster eine gewisse Validität und Gültigkeit (*valid*) aufweisen, sodass eine statistische Sicherheit für die Aussagen gegeben ist. Die Zusammenhänge zwischen den Daten sollen ebenso einen neuartigen also bislang unbekanntes (*novel*), potenziell nützlichen (*potentially useful*) und leicht verständlichen (*ultimately understandable*) Charakter aufweisen. Das im Rahmen des KDD entdeckte Wissen kann dazu beitragen, dass die gewünschten Wissensziele (z.B. in Bezug auf Kundenverhalten) erreicht sowie gegebenenfalls Maßnahmen festgelegt und durchgeführt werden. (Fayyad et al., 1996)

Die Einsatzgebiete des KDD sind kaum begrenzt und finden in vielen Bereichen, in denen Daten systematisch anfallen, eine breite Anwendung. Ein beliebtes

Anwendungsfeld ist beispielsweise das Marketing. Hier werden im Rahmen des KDD unter anderem Absatz- und Marktanteilsprognosen sowie Segmentierungen von Kunden in Verhaltensklassen durchgeführt. Weiterhin werden unter Verwendung des KDD Lieferantenbewertungen oder Bonitätsprognosen von Kunden im Bereich der Finanzierung sowie Früherkennungsanalysen von Unternehmensrisiken im Controlling realisiert. Andere Aufgabengebiete wie die Prognose von Fehlleistungen im Personal, optimale Standortwahl im Zuge des strategischen Managements oder die optimale Reihenfolgeplanung in der Produktion können ebenfalls mit dem KDD-Ansatz gelöst werden. (Wiedmann/Buckler, 2001)

Durch die Verwendung des KDD in vielen Einsatzbereichen im Unternehmen wird deutlich, dass KDD ein fachübergreifendes Forschungsgebiet darstellt und verschiedene Technologien sowie Methoden aus angrenzenden und sich überschneidenden Fachbereichen verwendet. Das KDD bedient sich aus den Forschungsrichtungen Statistik, Datenbanken und Maschinelles Lernen. Betont wird in diesem Zusammenhang, dass KDD mit den genannten Forschungsbereichen nicht gleichzusetzen ist. KDD unterscheidet und kennzeichnet sich besonders dadurch aus, dass sich die Ausrichtung des Wissensentdeckungsprozesses nicht auf eine bestimmte Anzahl an Phasen beschränkt sondern vielmehr den gesamten Entdeckungsprozess umschließt. (Fayyad et al., 1996)

3.2 Beispielm Modelle des KDD

Im Laufe der Zeit sind für das KDD verschiedene Modelle entwickelt worden, die den Entdeckungsprozess des Wissens unterschiedlich strukturieren und beschreiben. Das Stufenmodell nach Fayyad et al. (1996) sowie das Prozessmodell *Cross Industry Standard Process for Data Mining* (CRISP-DM) nach Chapman et al. (2000) gehören zu den gängigsten Vorgehensweisen, die sich insbesondere im Hinblick auf den Grad der Detaillierung und Ausrichtung unterscheiden. (Düsing, 2000)

Das in der Literatur häufig anzutreffende Stufenmodell nach Fayyad et al. (1996) strebt durch die Anwendung von verschiedenen Methoden an, hochwertiges Wissen aus Basisdaten zu extrahieren (Fayyad et al., 1996). Kennzeichnend für dieses Modell ist nicht der vollautomatisierte Ablauf, sondern die interaktive sowie iterative Prozessstruktur. Demzufolge müssen viele Entscheidungen basierend auf subjektiv bewerteten Informationen getroffen werden. Die angewendeten Methoden dienen demnach als Hilfestellung für den Anwender, der anhand der Zwischenergebnisse den weiteren Ablauf plant. Die Abbildung 5 gibt einen Überblick über die im KDD-Prozess zu durchlaufenden Stufen.

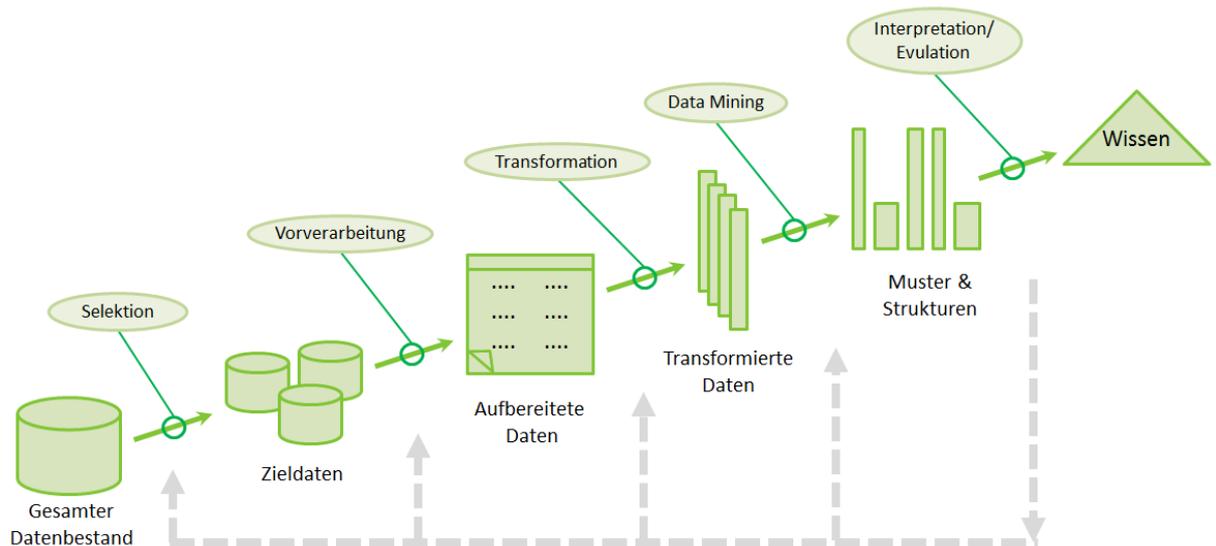


Abb. 5: KDD Stufenmodell (Fayyad et al., 1996)

Beginnend mit dem Business Understanding wird zunächst ein Verständnis für die Anwendungsdomäne entwickelt und das KDD-Ziel abgeleitet. Anschließend wird durch die Auswahl von zuvor bereinigten und vorbereiteten Datensätzen eine Analysebasis geschaffen. Dem folgt die Transformation der Daten, sodass im Anschluss geeignete Data Mining-Methoden für das bereits im Vorfeld definierte Ziel festgelegt werden. Schließlich wird eine Selektion von Modellen, Algorithmen und Parametern durchgeführt, mit deren Hilfe nach vorhandenen Datenmustern gesucht wird. Diese Datenmuster werden in einem nächsten Schritt interpretiert, sodass abschließend die erzielten Erkenntnisse in die Domäne zurück überführt werden können. (Fayyad et al., 1996)

Basierend auf Fayyad et al. (1996) werden im Folgenden die zu durchlaufenden Teilprozesse Domänenverständnis und Zieldefinition, Datenselektion, Datenvorbereitung und -bereinigung, Datentransformation, Data Mining-Methodenwahl, Algorithmen- und Hypothesenauswahl, Mustersuche durch Data Mining, Musterinterpretation sowie Wissensnutzung und -verbreitung detaillierter beschrieben. Bei der aufgeführten Beschreibung der Teilprozesse werden die drei Schritte Data Mining-Methodenwahl, Algorithmen- und Hypothesenauswahl und Mustersuche durch Data Mining unter Data Mining zusammengefasst werden, da hierzu im weiteren Verlauf der Kapitel näher eingegangen wird (Kapitel 3.3).

- **Domänenverständnis und Zieldefinition:** Zunächst wird von dem Anwender verlangt, sich ein Verständnis über die Domäne anzueignen. Hieran anknüpfend wird das Ziel der Analysen definiert, die sich aus der Sicht der gegebenen Anwendungen ergeben. Abhängig vom definierten Ziel, welches entweder die

Entdeckung von bestimmten Gegebenheiten oder die Prognose zukünftiger Aktivitäten beinhaltet, wird der weitere Prozessverlauf geplant und gesteuert.

- **Datenselektion:** Abhängig von dem benötigten Datenausschnitt, wird für den Selektionsprozess zunächst eine relevante Datenbasis bestimmt. Aus dieser Datenbasis heraus werden die für die Analyse erforderlichen spezifischen Daten aus der Rohdatenmenge selektiert. Hierbei wird diese Selektion im Hinblick auf die vorher definierten Ziele und dem angeeigneten Wissen über die Domäne durchgeführt.
- **Datenvorbereitung und -bereinigung:** Häufig beinhaltet der zuvor selektierte Datenausschnitt Fehler und Inkonsistenzen, die in diesem Teilprozess gesäubert und aufbereitet werden müssen. Durch die Säuberung der Daten wird eine Beseitigung und Korrektur falscher Einträge oder die Ergänzung von fehlenden Objekten beabsichtigt. Dieser Schritt der Erkennung und Beseitigung von Fehlern ist von besonderer Wichtigkeit, da Fehler die weiteren Analyseschritte und damit auch die Ergebnisse des gesamten KDD-Prozesses verfälschen können.
- **Datentransformation:** Im Rahmen der Datentransformation werden die in den vorherigen Schritten vorverarbeiteten Daten in eine zielgerichtete Form transformiert und so für die nachfolgenden Analyseschritte vorbereitet. Hierbei werden beispielsweise Datendimensionen reduziert und Transformationsmethoden angewendet, die zu einer Reduktion der betrachteten Anzahl von effektiven Variablen führen.
- **Data Mining:** Nachdem die Daten in den vorherigen Teilprozessen selektiert, bereinigt und transformiert wurden, wird im Schritt des Data Mining die eigentliche Datenanalyse durchgeführt. Hierbei wird ein besonderes Augenmerk auf das Erkennen von Mustern und Beziehungen in den transformierten Daten im Sinne des definierten Untersuchungsziels gelegt. Dies wird durch den Einsatz verschiedenster Data Mining-Methoden erreicht, wobei die Anwendung interaktiv und iterativ vom Anwender durchgeführt wird.

- **Musterinterpretation:** Die mit Hilfe des Data Mining-Prozesses ermittelten Ergebnisse werden zunächst mit den festgelegten Zielsetzungen verglichen und in Bezug auf diese interpretiert. Hierbei können sich neue Ziele und Fragestellungen entwickeln, die die einzelnen Prozessschritte nochmal wiederholen und im Anschluss analysiert werden müssen.
- **Wissensnutzung und -verbreitung:** Im letzten Schritt wird das neu entdeckte Wissen entweder direkt in das bestehende System einbezogen oder zunächst dokumentiert, sodass es an die Wissensinteressenten weitergeleitet werden kann. Dieser letzte Teilprozess kann ebenfalls als Überprüfung des bereits extrahierten Wissens oder als Lösung von potentiellen Konfliktsituationen im Unternehmen angenommen werden.

Der Einsatz von KDD-Modellen ist bedingt durch eine konsistente und qualitativ hochwertige Datenbasis. Diese Bereitstellung von Daten wird u.a. durch Data Warehouse-Konzepte zur Verfügung gestellt. Die Data Warehouse-Systeme führen weltweit verteilte Datenquellen in ein zentrales Informationsversorgungssystem zusammen. Hierdurch werden dann Daten für Anwendungs- und Analyseprogramme (wie z.B. für den KDD-Prozess) bereitgestellt. (Gabriel et al., 2009)

3.3 Data Mining als Phase im KDD-Prozess

Das Data Mining ist der wichtigste und anspruchsvollste Teilschritt des KDD-Prozesses und umfasst verschiedene Algorithmen und Methoden, die zur Wissensgewinnung beitragen (Kießwetter/Vahlkamp, 2007; Fockel, 2009). Bedingt durch die hohe Relevanz dieses Teilschritts, erfolgt im weiteren Verlauf der Arbeit zunächst eine generelle Beschreibung des Data Mining. Anknüpfend hierzu werden verschiedene Analysemethoden und deren Vorgehensweisen näher erläutert.

3.3.1 Generelle Beschreibung des Data Mining

Wie bereits erwähnt, entspricht Data Mining einem Teilschritt des KDD-Prozesses. Dieser besteht aus bestimmten Algorithmen, die durch Anwendung an einer vorgegebenen Datenbasis verschiedene Datenmuster und Auffälligkeiten erkennen. Der Ursprung der verwendeten Algorithmen liegt oftmals in der Mathematik, Statistik oder im Bereich der Künstlichen Intelligenz. (Fockel, 2009)

In der Literatur werden KDD und Data Mining häufig als Synonym für das gesamte Vorgehen verwendet (Fockel, 2009; Wiedmann et al., 2001). Allerdings sehen Fayyad et al. (1996) genau in diesem Vorgehen den Unterschied. KDD wird als der gesamte Prozess betrachtet, der für die Wissensentdeckung aus ausgewählten Daten zuständig ist, wobei im Hinblick auf das Data Mining lediglich nur ein Teil innerhalb des Gesamtprozesses verstanden wird. Somit werden beim Data Mining die restlichen Teilprozesse wie z.B. Datenselektion, Datenvorbereitung und -bereinigung, oder Datentransformation nicht mit einbezogen werden. Dabei besteht die Gefahr, dass ohne die restlichen Teilprozesse beim Data Mining bedeutungslose und ungültige Muster entdeckt werden können. (Fayyad et al., 1996)

Im Rahmen des Data Mining werden durch geeignete Methoden, die iterativ und wiederholt angewendet werden, transformierte Daten analysiert. Hierbei ergeben sich hauptsächlich zwei verschiedene Zielarten: die Verifikation und Entdeckung. Bei der Verifikation handelt es sich um die Überprüfung von zuvor aufgestellten Hypothesen, wohingegen bei der Entdeckung das Finden von neuen Datenmustern im Vordergrund steht. Die Entdeckung unterteilt sich wiederum in die Vorhersage von Zukunftswerten sowie in Beschreibung und Darstellung von Datenmustern. (Fayyad et al., 1996)

3.3.2 Gängige Methoden des Data Mining

Die Anwendungsbereiche des Data Mining werden in der Literatur nicht einheitlich beschrieben, jedoch kann man erkennen, dass sich im Wesentlichen vier Aufgabenfelder herausgebildet haben: Klassifikation, Segmentierung, Vorhersage und Assoziation. Durch diese vier Anwendungsbereiche sollen die oben beschriebenen Zielarten Verifikation und Entdeckung erreicht werden. (Chamoni et al., 2013; Nakhaeizadeh, 1999)

Die Aufgabe der Klassifikation besteht in der Identifizierung von Klassenzugehörigkeiten der gegebenen Objekte anhand verschiedener Merkmale, wobei das Ergebnis der Klassifikation zur Einteilung neuer Objekte in bereits vorhandene Klassen herangezogen wird. Ein weiterer Anwendungsbereich des Data Mining ist die Segmentierung. Dabei liegt das Ziel darin, ähnliche Objekte in neugebildete Gruppen einzuteilen. Hinsichtlich der Vorhersage werden unbekannte Merkmalswerte basierend auf anderen oder früheren Merkmalen prognostiziert. Hauptsächlich wird versucht, fehlende numerische Merkmalswerte zu ergänzen, wobei diese Merkmalswerte sich auf zukünftige Zeitpunkte beziehen. Zwischen der Vorhersage und der Klassifizierung gibt es einen entscheidenden Unterschied: bei der Vorhersage werden numerische Werte prognostiziert. Dahingegen handelt es sich bei der Klassifizierung um symbolische Werte. Den vierten Anwendungsbereich stellt die Assoziation dar. Als

grundlegende Aufgabenstellung wird hier das Finden von Modellen angesehen, die signifikante Zusammenhänge zwischen den Datenbasen beschreiben. (Fockel, 2009)

Tabelle 2: Anwendungsbereiche und Methoden des Data Mining (Bankhofer, 2004)

| Anwendungsbereich | Aufgabenstellung | Wesentliche Methoden |
|-----------------------|--|--|
| Segmentierung | Bildung von Gruppen aufgrund von Ähnlichkeiten der Objekte | <ul style="list-style-type: none"> • Clusteranalyse • Neuronale Netze |
| Klassifikation | Identifikation der Klassenzugehörigkeit von Objekten auf der Basis gegebener Merkmale | <ul style="list-style-type: none"> • Diskriminanzanalyse • Neuronale Netze • Entscheidungsbäume |
| Vorhersage | Prognose der Werte einer abhängigen kontinuierlichen Variable auf Basis einer funktionalen Beziehung | <ul style="list-style-type: none"> • Regressionsanalyse • Neuronale Netze • Entscheidungsbäume |
| Assoziation | Aufdeckung von strukturellen Zusammenhängen in Datenbasen mit Hilfe von Regeln | <ul style="list-style-type: none"> • Assoziationsanalyse |

Je nach den Anforderungen verschiedener Aufgabenbereiche, kann eine Vielzahl von unterschiedlichen Data Mining Methoden zum Einsatz kommen, die ihren Ursprung in der Mathematik, Statistik und künstliche Intelligenz haben (s. Tabelle 2). (Bankhofer, 2004)

Über die Tabelle 2 hinaus gibt es in der Literatur noch eine Vielzahl an weiteren Methoden, die jedoch im Rahmen dieser Arbeit nicht vollständig aufgezählt und erläutert werden. Vielmehr wird eine selektierte Auswahl an Data Mining-Methoden vorgestellt, welche im Kontext des SCMs eine verbreitete Anwendung erfahren. Dabei werden im konkreten folgende Methoden vorgestellt und erläutert: *Clusteranalysen, Künstliche Neuronale Netze, Entscheidungsbäume, Assoziationsanalysen.*

Clusteranalyse

Clusterverfahren werden verwendet, um die Datensätze in einer Gruppe derartig zusammenzufassen, dass die Datensätze innerhalb einer Gruppe möglichst homogen und Datensätze aus unterschiedlichen Gruppen möglichst heterogen sein sollen. Die Zugehörigkeiten der Datensätze werden durch das Ähnlichkeitsmaß bestimmt, welches die Grundlage für die Anwendung der Clusteranalyse darstellt. (Chamoni et al., 2013; Cleve/Lämmel, 2014)

Im Rahmen der Clusteranalyse gibt es ein breites Methodenspektrum an Algorithmen, die zur Gruppierung der Datensätze verwendet werden. Man kann hauptsächlich zwischen *partitionierenden* und *hierarchischen Verfahren* unterscheiden, wobei letztere sich in agglomerative und divisive Verfahren untergliedert. Aufbauend auf eine vorgegebene Gruppeneinteilung (Anzahl der Cluster) wird bei partitionierenden Verfahren versucht, die Gesamtlösung durch das iterative Austauschen der Datensätze zwischen den einzelnen Clustern zu verbessern. Im Gegensatz hierzu werden bei hierarchischen Verfahren keine vorliegenden Zerlegungen der Cluster optimiert, sondern es werden zunächst Cluster gebildet, die durch die Vereinigung mit anderen Clustern oder durch die Aufspaltung des vorliegenden Clusters schrittweise immer größere bzw. immer feinere Einteilungen erzeugen. (Chamoni et al., 2013)

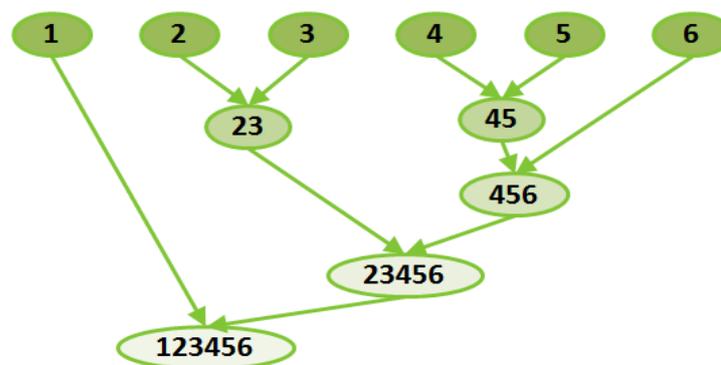


Abb. 6: Hierarchisches Clusterverfahren

Im Sinne der agglomerativ-hierarchischen Verfahren wird jeder Datensatz als ein eigenes Cluster aufgefasst. Hierauf basierend, werden diese einelementigen Cluster in mehreren Iterationsschritten nach und nach zusammengefasst. Im Gegensatz zu den agglomerativen Verfahren befinden sich bei divisiv-hierarchischen Verfahren bereits zu Beginn alle Datensätze in einem Cluster. Dabei wird bei jedem Iterationsschritt dieses Cluster fortlaufend in kleinere Untergruppen unterteilt. Allerdings findet dieses Verfahren in der Praxis selten Anwendung. (Chamoni et al., 2013)

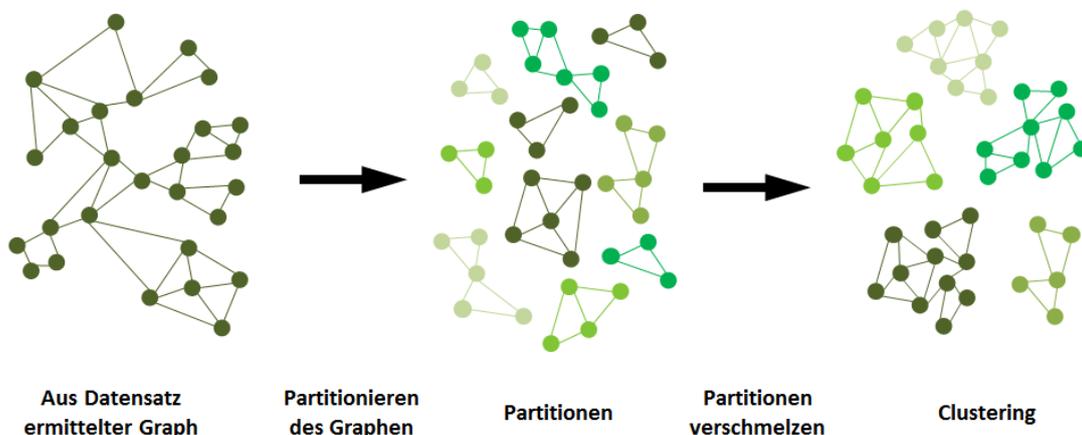


Abb. 7: Partitionierendes Clusterverfahren

Für das Data Mining sind besonders partitionierende Verfahren von Bedeutung, da hierarchische Verfahren aufgrund der exponentiell wachsenden Hierarchien bei großen Datenbeständen nicht optimal anzuwenden sind. Hinsichtlich der partitionierenden Clusteranalysen wird unabhängig von der Größe des Datenbestandes jeder Datensatz schrittweise in eine Gruppe eingeteilt, wobei die Einteilung basierend auf der Ähnlichkeit der Gruppe zum Datensatz erfolgt. (Bankhofer, 2004)

Künstliche Neuronale Netze

Künstliche Neuronale Netze sind informationsverarbeitende Systeme aus dem Forschungsgebiet der künstlichen Intelligenz. Dabei verfolgen Neuronale Netze das Ziel, gewisse Aufgaben des Data Mining zu bearbeiten indem eine Nachbildung der Arbeitsweise von Neuronen und deren Verknüpfungen im menschlichen Gehirn stattfindet. (Fockel, 2009)

Nach Wiedmann und Buckler (2001) ist der ausschlaggebende Vorteil hierbei, dass Neuronale Netze die folgenden drei Eigenschaften in sich vereinen:

1. **Nichtlinearität:** Es müssen weder ursächliche Zusammenhänge und Verknüpfungen zwischen den Variablen vorliegen noch sind sie auf lineare Beziehungen beschränkt. Jede beliebige Interaktion also nichtlineare Beziehung kann zwischen den Variablen nachgebildet werden.
2. **Lernfähigkeit:** Neuronale Netze sind lernfähig, d.h. dass die Zusammenhänge zwischen den Variablen selbstständig durch einen Lernprozess ermittelt werden können.
3. **Variablenanzahl:** Mit Neuronalen Netzen ist es möglich, während der oben genannten Eigenschaften eine Vielzahl an Variablen zu berücksichtigen.

Neuronale Netze sind von ihrer Konzeption her sehr flexible Systeme und werden bei der Erstellung von Klassifikations- und Regressionsmodellen und für die Clusterbildung erfolgreich eingesetzt (Fayyad et al., 1996). Der Einsatz ist besonders dann von Vorteil, wenn nicht bestimmte Wirkungszusammenhänge sondern vielmehr die Genauigkeit der Ergebnisse im Vordergrund steht (Backhaus et al., 2008).

Wiedmann und Buckler (2001) teilen die Ansicht, dass Neuronale Netz als Baukastensysteme interpretiert werden können, wobei ihre Bausteine, also die Neuronen, Grundrechenoperatoren darstellen. Hierbei besteht die Möglichkeit, alle logischen Zusammenhänge und Regeln durch eine geeignete Kombination der Bausteine abzubilden. Durch eine geeignete Zusammenschaltung und Verknüpfung der Neuronen entsteht schließlich ein Künstliches Neuronales Netz. Die nachfolgende

Abbildung 8 zeigt vereinfacht eine mögliche Darstellung eines Neurons mit Eingangs- und Ausgangskanten:

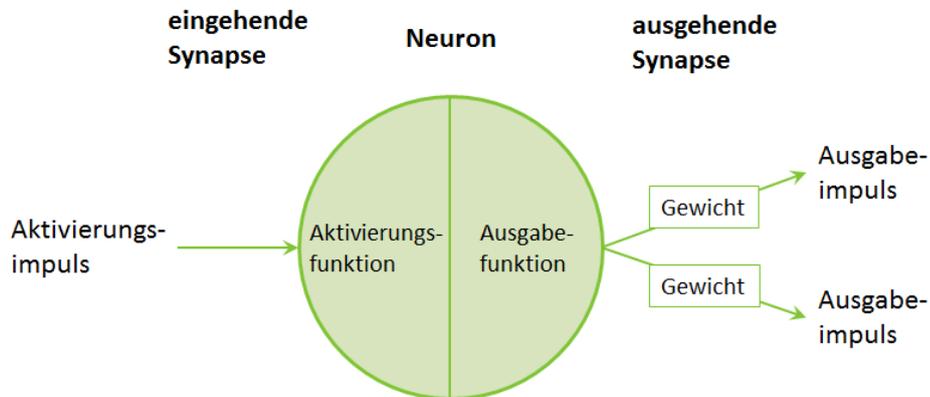


Abb. 8: Neuron (Hilpert, 2014)

Die Verarbeitung in einem Neuron wird in zwei Schritten durchgeführt. Zunächst wird überprüft, ob der Input einen gegebenen Schwellenwert übersteigt. Ist dies der Fall, wird das Neuron über eine Aktivierungsfunktion aktiviert, sodass es in einem nächsten Schritt über die Ausgabefunktion und gewichteten Verbindungen verarbeitete Signale an andere Neuronen weitergegeben werden können. Innerhalb des Neuronalen Netzes wird der Lernmechanismus durch das Anpassen der Gewichtungen in den ein- und ausgehenden synaptischen Verbindungen vollzogen. (Ertel, 2013)

Die Abbildung 9 zeigt die Grundstruktur eines zweischichtigen Neuronalen Netzes, welches aus einer Eingabe- und Ausgabeschicht besteht. In den verdeckten Schichten werden sowohl die Ausgabewerte des vorgelagerten Neuronen zusammengefasst als auch Transformationen durchgeführt, die nichtlinear sind. Somit wird ermöglicht, dass die Zusammenhänge der Ein- und Ausgabewerte durch das Neuronale Netz abgebildet werden können. (Fockel, 2009)

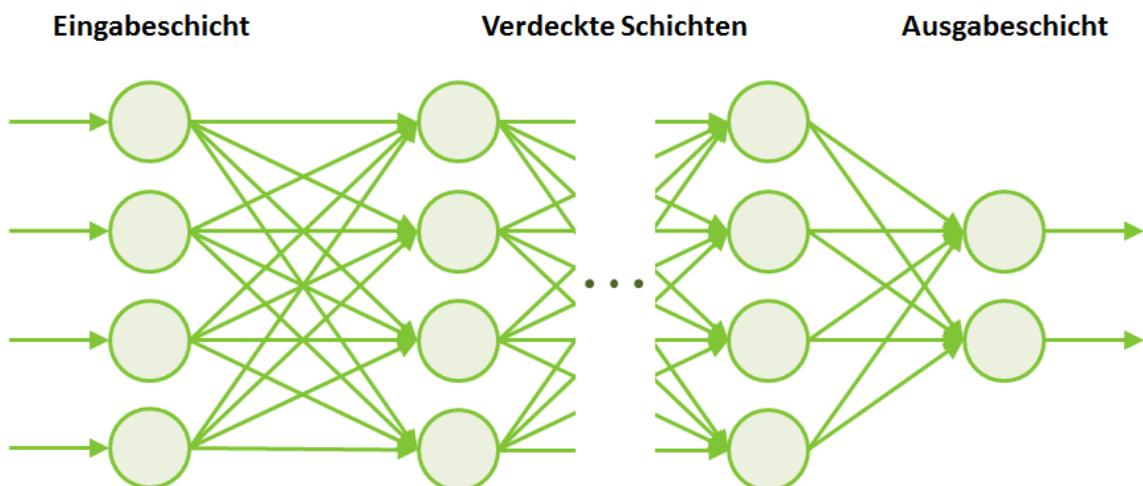


Abb. 9: Neuronales Netz mit variabler Anzahl verdeckter Schichten (Hilpert, 2014)

Durch die Vernetzung der Neuronen untereinander wird eine gemeinsame Informationsverarbeitung gewährleistet, dessen Richtung vorwärtsgerichtet (feedforward) oder rückwärtsgerichtet (feedbackward) sein kann. Während bei einem vorwärts gerichteten Netz die Informationsverarbeitung stets von der Eingabe- hin zur Ausgabeschicht erfolgt, können bei einem rückwärts gerichteten Netz sowohl Verbindungen zwischen den Neuronen einer Schicht existieren als auch die Informationsverarbeitung von nachgelagerten Neuronen ausgehend in Richtung der vorgelagerten Neuronen erfolgen. (Fockel, 2009)

Im Rahmen des Data Mining kommen jedoch vorwärts gerichtete Neuronale Netze, also Regressions- und Klassifikationsmodelle, häufiger zur Anwendung als rückwärts gerichtete. Mit Kohonen-Netzen, die auch als selbstorganisierende Karte bekannt sind, können Clusterbildungen vorgenommen werden und sind ebenfalls ein wichtiges Verfahren im Data Mining. (Chamoni et al., 2013)

Entscheidungsäume

Das Entscheidungsbaumverfahren ist ein gängiges und typisches Modell zur Klassifizierung von Datenbeständen und ist dadurch in vielen Data Mining-Werkzeugen aus dem Standardrepertoire nicht mehr wegzudenken. Das charakteristische Ziel einer Klassifizierung ist vor allem die Erzeugung von Modellen, die es ermöglichen ein Datenobjekt aus dem Datenbestand in mehrere, bereits vorgegebene Klassen einzuordnen. Dieses Vorgehen stellt auch die Grundidee der Entscheidungsäume dar, wobei der Datenbestand, die sogenannte Trainingsmenge, Schritt für Schritt in homogene Teilmengen aufgeteilt werden soll. Diese Schritte ermöglichen es, dass sich in einer Teilmenge mehr Datensätze einer Merkmalsausprägung befinden und diese somit einen höheren Homogenitätsgrad als die Ausgangsmenge aufweisen. Eines der zentralen Unterscheidungsmerkmale des Entscheidungsbaumverfahrens ist die Regel anhand derer die Aufteilung stattfindet. Zum einen gibt es klassische univariate Verfahren bei denen eine Aufteilung nur anhand eines Ausprägungsmerkmals vorgenommen wird. Zum anderen gibt es multivariate Verfahren, wohingegen auch eine Linearkombination der Ausprägungsmerkmale zur Aufteilung möglich ist. (Chamoni et al., 2013)

Ein Entscheidungsbaum besteht in seinem Aufbau aus mehreren verschiedenen Bestandteilen: den *Wurzelknoten*, den *Kanten*, den *inneren Knoten* und den *Blattknoten*. Der Wurzelknoten bezeichnet den Ausgangsknoten und umfasst somit alle dem Entscheidungsbaum zugrundeliegenden Daten. Im Gegensatz dazu werden im Durchlauf die inneren Knoten verwendet und bilden die jeweiligen Teilmengen ab. Als Endresultat eines Entscheidungsbaumes folgen die Blattknoten, die Endknoten des

Verfahrens und ordnen einem Objekt eindeutig eine Klasse zu. Soll nun ein neues Objekt anhand des Entscheidungsbaumes klassifiziert werden, wird zunächst am Wurzelknoten begonnen. Darauf folgend wird das Objekt über die Kanten so lange einem inneren Knoten zugeordnet, bis eine eindeutige Klassifizierung durch einen Blattknoten erfolgt ist. (Fockel, 2009)

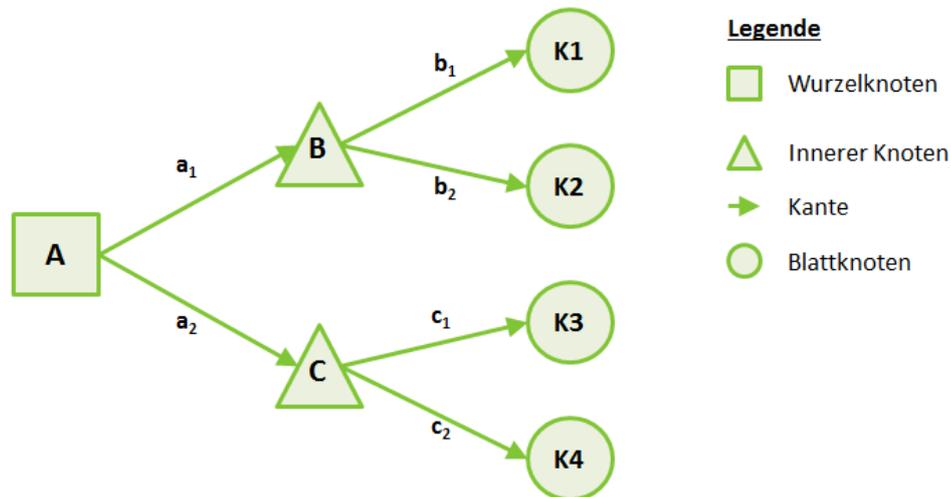


Abb. 10: Entscheidungsbaum (Hilpert, 2014)

Um die Qualität und Güte eines Entscheidungsbaumes zu bemessen wird die Fehlerklassifikationsquote eingesetzt. Sie wird durch den Quotienten der im Modell fehlerhaft klassifizierten Datensätze zur Gesamtzahl klassifizierter Datensätze ausgedrückt. Wichtig in diesem Zusammenhang ist jedoch, dass die Quote nicht anhand der Trainingsmenge sondern anhand einer Testmenge bestimmt wird. Zur Kontrolle kann jedoch auch die Quote für die Trainingsmenge bestimmt werden, welche aber eine geringe Aussagekraft hat, da sie durch weitere Aufteilung von Endknoten beeinflusst werden kann. Stellt man hierbei dann Fehlerklassifikationsquoten fest, die bei der Trainingsmenge klein und der Testmenge sehr hoch ausfallen, liegt ein sogenanntes Overfitting vor. Um diesem Effekt entgegenzuwirken werden einzelne Knoten sowie Kanten entfernt und somit die Quote auf Seiten der Testmenge verringert, aber auf Seiten der Trainingsmenge erhöht. (Chamoni et al., 2013)

Der Vorteil dieses Verfahrens liegt vor allem in der leicht verständlichen Darstellung der Bedingungen, die zur Bildung der Klassen genutzt werden. Weiterhin hat diese Darstellungsform auch zum Vorteil, dass das generierte Wissen leicht nachvollziehbar und schnell zugreifbar dargestellt ist. Allerdings können lange Rechenzeiten, die benötigt werden um die Entscheidungsbäume aufzustellen sich nachteilig auf den Prozess auswirken. Weiterhin kann bei großen Datenmengen die Struktur und Darstellung schnell unübersichtlich werden. Nichts desto trotz hat sich das Entscheidungsbaumverfahren als ein nichtwegzudenkendes und starkes Verfahren zur Klassifizierung von Datensätzen etabliert. (Chamoni et al., 2013; Ertel, 2013)

Assoziationsanalyse

Mit Hilfe der Assoziationsanalyse werden bestimmte Regeln erzeugt, die die Beziehungen zwischen sogenannten Items beschreiben, wobei Items die in den Datensätzen eines Datenbestandes enthaltenen Elemente darstellen und eine Transaktion eine Teilmenge von der Menge aller Items darstellt. Ein Beispiel für generierte Regeln ist die „wenn-dann“ Beziehung zwischen den Items: wenn Item A gilt, dann gilt auch Item B ($A \rightarrow B$). Für die Bewertung solcher Regeln sind zwei Maßzahlen von Bedeutung: der Support und die Konfidenz. Der Support einer Regel beschreibt mit welcher Wahrscheinlichkeit eine Regel im Gesamtdatenbestand auftritt und wird durch den Quotient aus der Anzahl der regelunterstützenden Transaktionen und der Anzahl aller Transaktionen ermittelt. Daraus kann man schließen, dass die Relevanz einer Regel umso größer ist, je höher der Supportwert liegt. Durch die zweite Maßzahl Konfidenz wird angegeben, mit welcher Wahrscheinlichkeit man eine Folgerung für eine bestimmte Ausprägung schließen kann, wenn die zu untersuchende Bedingung erfüllt ist. Außerdem ist bei der Generierung von Assoziationsregeln zu beachten, dass von den Regeln vorgegebene Mindestwerte zu überschreiten sind. Diese Mindestwerte werden im Rahmen des Supports (Mindestsupport) und der Konfidenz (Mindestkonfidenz) definiert, sodass Regeln, die unter diesen Schranken liegen, für die weitere Generierung von keinem großen Interesse sind. (Chamoni et al., 2013)

Eine Assoziationsanalyse läuft nach Agrawal et al. (1993) in folgenden zwei Schritten ab:

1. Zunächst werden alle Itemmengen bestimmt, die den Mindestsupportwert überschreiten. Diese Mengen stellen die *large itemsets* dar, wobei alle anderen als *small itemsets* definiert werden.
2. Anschließend werden zur Generierung der Regeln die zuvor ermittelten large itemsets verwendet und auf ihre Mindestkonfidenzwerte überprüft. Wenn dieser Wert ebenfalls überschritten wird, werden die Regeln mit ihren jeweiligen Support- und Konfidenzwerten ausgegeben.

Ein bekannter und häufiger Anwendungsfall der Assoziationsanalyse ist die Warenkorbanalyse. Dabei werden von Kunden zusammengestellten Warenkörbe daraufhin untersucht, welche Produkte bei einem Kauf kombiniert und gemeinsam erworben werden.

4 Einsatz von Data Mining Methoden im SCM

In diesem Kapitel werden die Möglichkeiten des Einsatzes von Data Mining-Methoden im Supply Chain Management evaluiert. Dabei bilden die Grundlagen von Supply Chains und des Wissensentdeckungsprozesses die theoretische Basis, auf der eine Gegenüberstellung von ausgewählten Fragestellungen im SCM (Kapitel 2.4) und den gängigen Data Mining-Methoden (Kapitel 3.3.2) erfolgt. Im Anschluss an die Gegenüberstellung werden die im SCM eingesetzten Data Mining-Verfahren tabellarisch aggregiert. Für eine effiziente Durchführung von Data Mining-Methoden wird eine hohe Datenqualität vorausgesetzt. Im Rahmen der nachfolgenden Analyseschritte dient ein vorverarbeiteter Datenbestand mit ebenfalls hoher Datenqualität als Grundlage, wobei sich das Vorgehen an dem in Kapitel 3.2 beschriebenen Modell von Fayyad et al. (1996) orientiert. Demzufolge wird vorausgesetzt, dass die Daten im Vorfeld gesichtet und selektiert werden sowie eine Vorverarbeitung und Transformation dieser vorgenommen wird, sodass im Anschluss die entsprechenden Data Mining-Methoden angewendet werden können.

4.1 Untersuchung der ausgewählten SCM-Fragestellungen und Data Mining-Methoden

Im Folgenden werden nun die relevanten Fragestellungen aus dem Bereich des SCMs durch die Anwendung verschiedener Data Mining-Methoden beantwortet. Die Analysen sind anhand der im SCOR-Modell vertretenen Prozesselemente strukturiert, welche zur Kategorisierung der Fragestellungen genutzt wurden (Kapitel 2.1 und Kapitel 2.4). Bei der Darstellung der jeweiligen Fragestellung werden zunächst ihre Thematik und die daraus resultierenden Problemfelder erörtert. Weiterhin werden exemplarisch mögliche Zuordnungen von Anwendungsbereichen des Data Mining vorgenommen (Kapitel 3.3.2) und es wird dargestellt, welchen Nutzen die aus den großen Datenbeständen gezogenen Kenntnisse bringen können.

1) Wie hoch werden zukünftige Kundenbedarfe sein?

Die Bedarfsplanung ist entscheidend für alle weiteren Planungsschritte in der Wertschöpfungskette und ist im Rahmen des SCM-Aufgabenmodells der taktischen und operativen Planungsebene, also dem Supply Chain Planning, zuzuordnen. Die Bedarfsermittlungen werden zum größten Teil in kurz-, mittel- und langfristig anhaltende Bedarfsverhalten der Kunden unterschieden und können einen konstanten, saisonalen sowie trendhaften Charakter aufweisen. Ausgehend von dem Bedarfsverhalten der Kunden können die verschiedenen Unternehmensbereiche wie

Produktion oder Beschaffung umso besser planen oder einkaufen, je genauer der Bedarf im Vorfeld ermittelt wurde. Es wird somit eine bedarfsgerechte Produktion und Bereitstellung der angeforderten Kundenwünsche angestrebt (Kapitel 2.2). Die Bestimmung des Kundenbedarfs kann durch verschiedene Prognosemethoden des Data Mining erfolgen. Demzufolge kann diese Fragestellung dem Anwendungsbereich der Vorhersage zugeordnet werden, welche die Werte einer abhängigen sowie kontinuierlichen Variablen auf Basis einer funktionalen Beziehung prognostiziert (Kapitel 3.3.2). Die Prognose von unbekanntem Merkmalswerten, also hier von zukünftigen Kundenbedarfen, kann auf der Basis anderer Merkmale, wie z.B. dem Abgleich von vorhandenen Produktionskapazitäten und Beständen, oder von Werten des gleichen Merkmals aus früheren Perioden durchgeführt werden. Das Ergebnis solcher Prognosen ist dann in diesem Fall der erwartete Bedarfsverlauf für das jeweilige zu prognostizierende Produkt.

2) Wo entsteht der Bedarf? Bei welchen Kunden bzw. in welchen Regionen liegt dieser Bedarf vor?

Wie bereits in Frage 1) erwähnt, hat die Kundenbedarfsplanung in vielen Unternehmensbereichen des SCMs eine entscheidende Rolle (Kapitel 2.2). Aufgrund einer gesamten Betrachtung des Kundenverhaltens können in der Planungsperiode wichtige Schlussfolgerungen gezogen werden. Ergänzend zu den im Rahmen des Data Mining prognostizierten Werten aus Frage 1), kann die Zuordnung der ermittelten Bedarfe in einzelne Regionen von hoher Relevanz sein. Infolgedessen kann durch eine regionale Anpassung ein kundenspezifisches Sortiment für die einzelnen Regionen entlang der Wertschöpfungskette geplant werden, der ebenfalls zur Steigerung der Kundenbindung beitragen kann. Hinsichtlich des Data Mining kann durch die Anwendung von Segmentierungsmethoden eine Bedarfszuordnung erfolgen (Kapitel 3.3.2). Mit Hilfe der Bildung von Clustern können die Bedarfe aus ähnlichen Regionen zusammengefasst werden, sodass aus diesen Erkenntnissen heraus bestimmt werden kann, in welchen Regionen hohe oder niedrige Kundenbedarfe vorliegen. Dies kann zu Folge haben, dass durch eine effiziente Planung der in den verschiedenen Unternehmensstandorten vorliegenden Produktionsauslastungen u.a. geringere Lagerbestände erreicht werden können.

3) Welche Distributionszentren sollen wo eingerichtet werden und welche Kunden beliefern?

Eines der wichtigsten Elemente einer Supply Chain sind die Struktur und der Aufbau von vorhandenen Distributionsnetzwerken. Dabei können unterschiedlichste Formen auftreten: vom direkten Beliefern über die Bündelung von Lieferungen bis hin zu Distributionslagern und -zentren vor Ort bei den Kunden. Die Distribution ist im SCM

dem Supply Chain Planning, also dem taktischen und operativen Aufgabenbereich, zuzuordnen und betrachtet essentielle Faktoren, wie z.B. die Beachtung von Transportkosten und schnelle Auslieferungszeiten (Kapitel 2.2). Ein Aufbau von Distributionslagern in zentralen Gebieten, in denen eine hohe Konzentration der Kundendichte vorliegt, kann für eine effiziente Struktur von enormem Vorteil sein. Dieser Vorteil kann in der Vermeidung von häufigen und kostenintensive Materialbewegungen sowie langen Transportwegen liegen. Dadurch kann eine Steigerung der Warenverfügbarkeit an wichtigen Umschlagpunkten der Distribution erzielt werden. Für die Entscheidung über die optimale Lage von Distributionszentren kann hinsichtlich der Data Mining-Methoden auf Anwendungsbereich der Segmentierung zurückgegriffen werden (Kapitel 3.3.2). Durch Bildung von Clustern ähnlicher Objekte, also hier ähnlicher Kundenstandorte, kann im Clusterzentrum die Lage des Distributionszentrums gesetzt werden. Wichtig zu erwähnen ist, dass bei der Entscheidung weitere Faktoren (z.B. Kosten sowie Verfügbarkeit von Kapazitäten) zu beachten sind.

4) Bei welchem Lieferanten soll bestellt werden?

Neben den Herstellern, Distributoren, Dienstleistern und Kunden gehören die Lieferanten zu den wichtigsten Akteuren in einer Supply Chain. Die durchschrittenen Entwicklungsstufen zeigen, dass sich der Informationsaustausch zwischen Kunden, Lieferanten und Dienstleistern im Laufe der Zeit intensiviert und somit einen immer höheren Stellenwert erreicht hat (Kapitel 2.1). Durch eine Verbesserung der eingehenden Lieferantenströme, kann es dem Unternehmen gelingen jegliche Prozesskosten zu reduzieren, Produktqualitäten zu erhöhen sowie die Beziehungen zu ihren Lieferanten zu optimieren. Für einen erfolgreichen Lieferantenstrom ist es daher unabdingbar eine geeignete Lieferantanalyse und -bewertung durchzuführen. Diese Bewertungen werden im Rahmen der Beschaffungsplanung vorgenommen, welche ein Teil der im SCM-Aufgabenmodell aufgeführten operativen und taktischen Planungsebene darstellt (Kapitel 2.2). Ausgehend von den im Vorfeld durchgeführten Bedarfsermittlungen, können für die Festlegung einer geeigneten Lieferquelle die Ermittlung, Aufbereitung, Verarbeitung und Darstellung verschiedener Informationen über die potentiellen Lieferanten eine wichtige Rolle spielen. Für die Gewinnung solcher Informationen kann im KDD-Prozess die Data Mining-Methode der Klassifikation herangezogen werden (Kapitel 3.3.2). Hierbei kann der Lieferantenauswahlprozess in verschiedene Teilschritte, wie z.B. die Lieferantenidentifikation, Lieferanteneingrenzung, Lieferantenbewertung und Lieferantenverhandlung, mit entsprechenden Filtergrößen untergliedert werden. In dem ersten Teilschritt der Lieferantenidentifikation werden hinsichtlich der Klassifikation verschiedene Filtergrößen, wie Branche, Produktspektrum oder Verfahrensvarianten als

Klassen betrachtet, denen anschließend diverse Lieferanten als Objekte zugeordnet werden. Auf diese Art kann ebenfalls in den weiteren Teilschritten des Lieferantenauswahlprozesses, verschiedenen Klassen die zugehörigen Objekte zugeordnet werden, sodass anhand von notwendigen Auswahlkriterien die optimale Lieferantenquelle selektiert werden kann.

5) Welche Produkte sollen in welchen Mengen beschafft werden, um eine optimale Lagerverwaltung und Materialflüsse zu gewährleisten?

Die Beschaffungsplanung ist im Rahmen des SCM-Aufgabenmodells der taktischen und operativen Planungsebene zuzuordnen und dient zur mengen- und termingerechten Versorgung der Produktions- und Lagerstandorte mit Produkten sowie den Materialien, die für die Herstellung dieser Produkte notwendig sind (Kapitel 2.2). Für die Optimierung der Lagerverwaltung können verschiedene Anwendungsbereiche des Data Mining verwendet werden. Eine Möglichkeit besteht in der Klassifikation (Kapitel 3.3.2), indem z.B. Materialien oder Aufträge auf Basis gegebener Merkmale in Klassen zugeordnet werden. Hierbei können verschiedene primäre Ziele, wie z.B. eine kostenoptimierte Lagerhaltung und niedrigere Ausbringungszeiten, definiert werden. Eine weitere Möglichkeit bietet sich durch die Anwendung der Assoziationsanalyse an (Kapitel 3.3.2). Hinsichtlich der Verwaltung von Materialflüssen können die mit Hilfe der Assoziationsregeln aufgestellten Wenn-Dann-Beziehungen zu zusammengeführten Aufträgen führen, deren Lieferorte sich z.B. in unmittelbarer Nähe befinden und somit zu Weg- und Zeiteinsparungen führen. Der Anwendungsbereich der Segmentierung kann hinsichtlich der Lagerverwaltung ebenfalls zum Einsatz kommen (Kapitel 3.3.2). Hierbei ist es möglich die im Lager zu kommissionierenden Aufträge nach Produktstandorten zusammenzufassen, um so die Kommissionierzeiten zu optimieren. Demnach können Materialien, die nach und nach an dem gleichen Produktionsstandort einzusetzen sind, an einem gleichen Lagerstandort aufbewahrt werden.

6) Wie hoch ist die Qualität der produzierten Teile?

Die Qualität der hergestellten Güter stellt einen wichtigen Wettbewerbsfaktor dar und leistet einen entscheidenden Beitrag für das Erfolgspotential der Supply Chain. Sie kann anhand von Kriterien wie Ausschuss- und Nacharbeitsraten während des Produktionsprozesses oder auch an der Anzahl der Rücklieferungen durch den Kunden bewertet werden. Durch die Vermeidung der genannten Kriterien wird eine Befriedigung der Kundenanforderungen gefordert, sodass die Unternehmen in einer Supply Chain eine fortlaufende Qualitätsverbesserung und -einhaltung anstreben (Kapitel 2.2). Dieses entscheidende Kaufkriterium kann jedoch nur dann erreicht werden, wenn die Unternehmensprozesse dazu geeignet sind die kundenspezifischen Wünsche zu erfüllen. Eine Nichteinhaltung der Qualitätsanforderungen kann

verschiedene Auswirkungen haben. Zum einen können fehlerhafte Produkte oder Produktionsrückrufe den Gewinn der Unternehmung begrenzen und zum anderen können sie dem Unternehmensimage schaden. Im Rahmen des Data Mining kann beispielsweise die Qualität von Bauteilen mit Hilfe von Klassifikationsmethoden identifiziert und optimiert werden. Damit die versteckten Produktfehler frühzeitig erkannt werden können, ist eine gezielte Analyse von Produktdatenbeständen hinsichtlich der Qualitätsabweichungen erforderlich (Kapitel 3.2.2). Hierbei kann die Klassifikation jeden Datensatz, in diesem Fall ein Bauteil, separat z.B. in gute und schlechte Bauteile bewerten. Um eine solche Methode durchführen zu können, werden sowohl historische Daten und zusätzlich noch die Bewertung durch den Anwender benötigt. Der Anwender bewertet dann, ob die betrachteten historischen Datensätze jeweils mit gut, mittel oder schlecht klassifiziert werden sollen. Basierend auf dieser Zuordnung, werden mit der Klassifikationsmethodik Ähnlichkeiten und Unterschiede zwischen den Datensätzen gesucht, sodass im Anschluss für neue Datensätze die richtigen Klassen vorausgesagt werden können. Allerdings kann hierbei die Genauigkeit von den verfügbaren Messwerten sowie deren Menge und Qualität abhängen.

7) Wie oft müssen Wartungen an Produktionsanlagen durchgeführt werden um zukünftige Maschinenausfälle zu vermeiden?

Der Herstellungsprozess ist in einer Wertschöpfungskette von hoher Bedeutung, wobei besonders im Hinblick auf Maschinenausfälle die vorbeugende Wartung der Maschinen eine wichtige Rolle einnimmt (Kapitel 2.4). Aufgrund von unerwarteten Ausfällen der in die Fertigungskette integrierten Produktionsanlagen, kann ein Stillstand der gesamten Produktion eintreten. Daher können durch vorbeugende Wartungsmaßnahmen u.a. Maschinenstillstände aufgrund sich verändernder Prozessparametern vermieden werden. Für die Gewährleistung der Verfügbarkeit von Produktionsanlagen können im Rahmen des Data Mining verschiedene Anwendungsbereiche eingesetzt werden. Durch diesen Einsatz von DM-Methoden wird eine gezielte Auswertung der erfassten Produktionsdaten ermöglicht, sodass frühzeitig Maschinenstillstände, Verzögerungen oder unerwartete Ereignisse prognostiziert werden können. Für die Unterstützung der Instandhaltung kann z.B. der Anwendungsbereich der Vorhersage herangezogen werden, sodass basierend auf Vergangenheitsdaten durch die Prognose die Notwendigkeit einer Wartung vorausgesehen wird (Kapitel 3.3.2). Weiterhin können anhand von Entscheidungsbäumen relevante Attribute, wie z.B. die Laufzeit oder die Leistung, also die hergestellte Stückzahl, der Anlage bestimmt werden, sodass anschließend der optimale Wartungszeitpunkt vorhergesagt wird. Auf diese Weise kann die Instandhaltungsplanung gezielte Hinweise auf durchzuführende Wartungsarbeiten erhalten und somit sehr präzise Wartungspläne erstellen. Basierend auf den

Erkenntnissen können die Lebensdauer von Anlagen, die Ersatzteilversorgung und die Betriebsabläufe optimiert werden.

8) Wie lassen sich die Rüstzeiten optimieren?

Bereits in Kapitel 2.2 werden die verschiedenen Teilziele im SCM aufgeführt, wobei u.a. die Verkürzung der Prozess- und Durchlaufzeiten aufgeführt. Mit der Durchlaufzeitverkürzung einhergehend wird ebenfalls eine auftragsbezogene Rüstzeitoptimierung angestrebt. Entscheidend für die Reduzierung der Umrüstvorgänge eines Produktionssystems ist die herzustellende Losgröße. Um Produktionsaufträge mit ähnlichen Produktionsschritten und Rüstoperationen zusammenzufassen und somit die Rüstzeiten zu senken, können hinsichtlich des Data Mining verschiedene Anwendungsbereiche herangezogen werden. Im Rahmen der Segmentierung können beispielsweise Cluster ähnlicher Objekte (Kapitel 3.3.2), hier Produktionsgüter, gebildet werden, die während eines Produktionsslots gefertigt werden und untereinander keine oder nur geringe Unterschiede in den Rüstzeiten aufweisen. Anhand dieser Gruppierungen können neue Kennzahlen, wie z.B. Verzögerungszeiten, erfasst und deren Ursachen analysiert werden. So können in Losgrößen- und Reihenfolgeplanungen Lösungen erarbeitet werden, die zu einer fertigungsgerechten Bearbeitungsreihenfolge mit optimierten Rüstzeiten führen. Eine weitere Möglichkeit zur Optimierung des Produktionsprogramms hinsichtlich der Rüstzeiten bietet der Anwendungsbereich der Assoziation an. Hierbei werden mit Hilfe von Regeln strukturelle Zusammenhänge in großen Datenbeständen aufgedeckt (Kapitel 3.3.2). Mit Hilfe der in der Assoziationsanalyse gewonnenen Erkenntnis, welche Produkte häufig gemeinsam oder zeitnah herzustellen sind, können so Produktionsreihenfolgen angepasst und infolgedessen Rüstzeiten und Lagerbestände gesenkt werden.

9) Welches Transportmittel oder welcher Transportweg ist wann am effizientesten?

Aus dem bereits in Kapitel 2.2 vorgestellten SCM-Aufgabenmodell geht hervor, dass bei der Distributionsplanung vor allem konkrete Kundenaufträge aus dem Auftragsmanagement, prognostizierte Transportbedarfe aus der Netzwerkplanung sowie Ergebnisse aus der Produktionsfeinplanung zu berücksichtigen sind. Dabei kann die richtige Wahl geeigneter Transportmittel und -wege einen relevanten Faktor in der Supply Chain darstellen. In diesem Zusammenhang ist nicht nur eine hohe Kundenzufriedenheit durch kurze Lieferzeiten möglich, sondern auch bei einer optimalen Nutzung und Kombination der Liefermöglichkeiten hohe Kostenersparnisse zu erzielen. Hinsichtlich der Wahl von optimalen Transportmitteln und -wege kann im Rahmen der Data Mining-Werkzeuge der Anwendungsbereich der Klassifikation als auch die der Segmentierung zum Einsatz kommen. Bei der Festlegung der Transportmittel bieten sich die Klassifikationsmethoden besonders gut für die

Anwendung an (Kapitel 3.3.2). Hierbei kann eine Zuordnung von Transportmitteln zu den Strecken des Distributionsnetzwerkes stattfinden. Weiterhin können durch die Festlegung der Haupt-Transportrouten bestimmte Regionen des Distributionsnetzwerkes den Gebietsspediteuren zugeteilt werden. Im Vorfeld sollte jedoch zunächst das Sendungsaufkommen bezüglich der einzelnen Regionen erfasst werden, wobei der Anwendungsbereich der Segmentierung herangezogen werden kann. Hierbei können die Sendungsaufkommen gleicher oder ähnlicher Regionen gruppiert und so geeignete Lieferstrecken für das Netzwerk (z.B. bezogen auf festgelegte Zeiteinheiten) bestimmt werden.

10) Wie kann man die Anzahl der verspäteten Aufträge minimieren und eine fristgemäße Lieferung sicherstellen?

Die Liefertreue zu einem seitens der Unternehmung bestätigten Termin ist ein wichtiger Faktor im Rahmen des SCMs. Sowohl eine verspätete als auch eine verfrühte Auslieferung von Produkten kann den Kunden zum Teil vor enorme Schwierigkeiten stellen: beispielweise kann eine zu frühe Lieferung dazu führen, dass kurzfristig ein Lagerplatz gefunden werden muss, da noch keine freien Lagerplätze zur Verfügung stehen. Hinsichtlich der Lieferung von Aufträgen umfasst dieser Prozess u.a. auch die Steuerung und Überwachung der Transportprozesse (Kapitel 2.1). In diesem Sinne sollten ebenfalls die Tourenplanungen der Transportmittel darauf ausgerichtet sein, neben anderen Faktoren (z.B. Kapazitätsauslastungen der Transportmittel) auch eine fristgemäße Lieferung sicherzustellen. In Bezug auf die im KDD-Prozess eingesetzten Data Mining-Methoden eignet sich bezüglich der Fragestellung der Anwendungsbereich der Segmentierung besonders gut. Durch die Segmentierung von Datensätzen ist es möglich Gruppen von Objekten zu bilden, die untereinander eine gewisse Ähnlichkeit aufweisen (Kapitel 3.3.2). Dementsprechend kann hinsichtlich der Fragestellung eine Gruppe von verspäteten Aufträgen ausfindig gemacht werden. Basierend auf dieser Gruppe von Datensätzen können Merkmale identifiziert werden, die für die Verspätungen von hoher Bedeutung oder möglicherweise sogar ursächlich sind. Um eine zuverlässige sowie termingerechte Lieferung der Aufträge zu gewährleisten und somit die Anzahl der verspäteten Aufträge zu senken, können verschiedene Kennzahlen festgelegt werden. Mit Hilfe dieser Kennzahlen kann bestimmt werden, an welchen Stellen der Wertschöpfungskette ein Eingreifen durch die Planung notwendig ist, um frühzeitig den drohenden Verspätungen von Aufträgen entgegenzuwirken und Maßnahmen zu ergreifen.

11) Worauf lässt sich der Defekt bzw. die Beschwerde am rückgelieferten Produkt zurückführen?

Die Abwicklung sowie die Prüfung von Kundenbeschwerden bzw. rückgesendeten Produkten mit Defekten sind wichtige Komponenten des SCOR-Prozesselements Rückliefern (Kapitel 2.1). Aus diesen Kundenbeschwerden eröffnet sich den Unternehmen die Möglichkeit, Verbesserungspotenziale an ihren Produkten, Dienstleistungen und Prozessen zu identifizieren und diese in der Praxis zu realisieren. Es lässt sich also leicht schlussfolgern, dass die Bearbeitung von den Beschwerden und Rücklieferungen nicht ohne weiteres unbeachtet bleiben können. Vielmehr ist es ein bestrebenwerteres Ziel systematisch den Bearbeitungsprozess einer Beschwerde bis hin zu ihrer Lösung zu verfolgen. Beschaffungsseitig als auch absatzseitig sollten Retouren-Prozesse berücksichtigt werden, da unter Vernachlässigung dieser Aspekte Image und damit einhergehend Gewinnmargen des Unternehmens leiden und einbrechen können. Vor allem um Kosten und Verluste zu vermeiden ist es notwendig, Informationen und Daten des Beschwerdemanagements zu analysieren. Hierbei kann die Anwendung von Data Mining-Methoden genutzt werden. Hinsichtlich der Segmentierungsmethoden kann anhand der vorliegenden Datenmengen von rückgelieferten Produkten untersucht werden, welche Merkmale die Daten aufweisen (Kapitel 3.3.2). Dazu kann eine Gruppe von rückgelieferten Aufträgen zusammengeführt werden, die gleiche oder ähnliche Merkmale, also gleiche oder ähnliche Gründe für ihre Rücklieferung, aufweisen. Basierend auf diesen Merkmalen kann die Ursache für die Rücklieferung analysiert und schließlich behoben werden. Des Weiteren bietet sich ebenfalls der Einsatz von Assoziationsanalysen an, wobei durch die Aufstellung von „wenn-dann“ Beziehungen strukturelle Zusammenhänge aufgedeckt werden können (Kapitel 3.3.2). Bezogen auf die Fragestellung, kann eine solche Beziehung folgendermaßen lauten: wenn Bauteil A eingebaut wird, dann erfolgt nach „X“ Wochen eine Produktbeschwerde. Somit ist es möglich, Mängel oder Fehler an der Produktionskette zu identifizieren und zu eliminieren.

4.2 Systematische Gegenüberstellungen der eingesetzten Data Mining-Verfahren im SCM

Ziel dieses Kapitels ist eine tabellarische Gegenüberstellung der im Rahmen des SCM beispielhaft aufgestellten Fragestellungen und den gängigen Data Mining-Methoden, mit der Intention eine kompakte Übersicht für Unternehmungen bereit zu stellen. Dazu werden nun die textuellen Ergebnisse aus dem vorherigen Kapitel 4.1 in Tabelle 3 zusammengeführt.

Tabelle 3: Gegenüberstellung von SCM-Fragen und Data Mining Anwendungsbereichen

| Pos. | Aktivität | Fragestellung aus dem SCM | Anwendungsbereich im Data Mining |
|---------------|--------------------------------------|--|---|
| Planen | | | |
| 1 | Bedarfsplanung | Wie hoch werden zukünftige Kundenbedarfe sein? | <p>Vorhersage:</p> <ul style="list-style-type: none"> • Bedarf wird anhand von Merkmalen wie z.B. gesammelten Informationen über bisherige Einkäufe prognostiziert • Visuelle Darstellung der Resultate anhand von Verlaufskurven |
| 2 | Zuordnung und Verteilung der Bedarfe | Wo entsteht der Bedarf? Bei welchen Kunden bzw. in welchen Regionen liegt dieser Bedarf vor? | <p>Vorhersage:</p> <ul style="list-style-type: none"> • Ermittlung des regionalen Kundenbedarfs <p>Segmentierung:</p> <ul style="list-style-type: none"> • Gruppenbildung von Bestellungen aus ähnlicher Region führt zu Bedarfsgruppenerkennung → Spezifische Produktplanung wird ermöglicht |
| 3 | Netzwerkplanung | Welche Distributionszentren sollen wo eingerichtet werden und welche Kunden beliefern? | <p>Segmentierung:</p> <ul style="list-style-type: none"> • Bildung von Gruppen ermöglicht erste Bestimmungen zur Lage von Distributionszentren |

| Pos. | Aktivität | Fragestellung aus dem SCM | Anwendungsbereich im Data Mining |
|-------------------|--------------------------------|--|--|
| Beschaffen | | | |
| 4 | Lieferantenbewertung | Bei welchem Lieferanten soll bestellt werden? | <p>Klassifikation:</p> <ul style="list-style-type: none"> • Zuordnung von Lieferanten in Klassen mit spezifischen Objekteigenschaften → durch die Auswahl benötigter Eigenschaften kann ein bestimmter Lieferant ermittelt werden |
| 5 | Bestellung für Bedarf absetzen | Welche Produkte sollen in welchen Mengen beschafft werden, um eine optimale Lagerverwaltung und Materialflüsse zu gewährleisten? | <p>Segmentierung:</p> <ul style="list-style-type: none"> • Gruppenbildung von im Lager zu kommissionierenden Materialien nach Produktstandorten führt zu einer Optimierung der Kommissionierzeiten <p>Klassifikation:</p> <ul style="list-style-type: none"> • Zuordnung von Materialien in vorgegebene Klassen mit verschiedenen Zielsetzungen <p>Assoziation:</p> <ul style="list-style-type: none"> • Durch „wenn-dann“-Beziehung optimierte Verwaltung von Materialflüssen |

| Pos. | Aktivität | Fragestellung aus dem SCM | Anwendungsbereich im Data Mining |
|-------------------|---------------------------|--|---|
| Herstellen | | | |
| 6 | Qualitätskontrolle | Wie hoch ist die Qualität der produzierten Teile? | Klassifikation: <ul style="list-style-type: none"> • Klassifizierung und Bewertung von guten und schlechten Produkten kann zur Steigerung der Qualität führen |
| 7 | Instandhaltungsplanung | Wie oft müssen Wartungen an Produktionsanlagen durchgeführt werden um zukünftige Maschinenausfälle zu vermeiden? | Vorhersage: <ul style="list-style-type: none"> • Frühzeitiges Vorhersagen von Maschinenausfällen und Wartungen, sodass Verzögerungen im Betriebsablauf reduziert werden |
| 8 | Produktionsprozess planen | Wie lassen sich die Rüstzeiten optimieren? | Segmentierung: <ul style="list-style-type: none"> • Gruppenbildung von Produktionsgüter mit ähnlichen Rüstzeiten führt zu einer Analyse ihrer Ursache und zur Rüstzeitoptimierung Assoziation: <ul style="list-style-type: none"> • Aufdecken von strukturellen Zusammenhängen anhand von „wenn-dann“-Beziehungen führt zu der Erkenntnis, welche Produkte gemeinsam oder zeitnah herzustellen sind |

| Pos. | Aktivität | Fragestellung aus dem SCM | Anwendungsbereich im Data Mining |
|--------------------|---------------------------|---|---|
| Liefern | | | |
| 9 | Auftrag transportieren | Welches Transportmittel oder welcher Transportweg ist wann am effizientesten? | <p>Segmentierung:</p> <ul style="list-style-type: none"> Gruppenbildung von Sendungsaufkommen aus ähnlichen Regionen führt zu einer Bestimmung von geeigneten Lieferstrecken <p>Klassifikation:</p> <ul style="list-style-type: none"> Zuordnung von Transportmitteln zu den Strecken im Distributionsnetzwerk |
| 10 | Liefertreue sicherstellen | Wie kann man die Anzahl der verspäteten Aufträge minimieren und eine fristgemäße Lieferung sicherstellen? | <p>Segmentierung:</p> <ul style="list-style-type: none"> Gruppenbildung von verspäteten Aufträgen mit ähnlichen Merkmalen führt zur Erkennung der Ursache für Verspätungen |
| Rückliefern | | | |
| 11 | Rücklieferung prüfen | Worauf lässt sich der Defekt bzw. die Beschwerde am rückgelieferten Produkt zurückführen? | <p>Segmentierung:</p> <ul style="list-style-type: none"> Gruppenbildung von rückgelieferten Produkten mit ähnlichen Defekten führt zur Erkennung der Gründe für die Rücklieferung <p>Assoziation:</p> <ul style="list-style-type: none"> Aufdecken von strukturellen Zusammenhängen anhand von „wenn-dann“-Beziehungen führt zu einer Fehleridentifikation in der Herstellung |

Aus Tabelle 3 wird ersichtlich, dass hinsichtlich des Data Mining besonders der Anwendungsbereich der Segmentierung häufig zur Beantwortung der Fragen eingesetzt werden kann. Dies lässt sich dadurch begründen, dass die Segmentierung in sehr vielen Anwendungsfeldern Einsatz finden kann: von der Kundensegmentierung, über die Standortplanung bis hin zur Produktentwicklung.

5 Exemplarische Fallbeispiele

Im Folgenden stehen vereinfacht exemplarische Fallbeispiele im Vordergrund, wobei alle zur Durchführung einer Datenanalyse notwendigen Stufen näher betrachtet werden. Als ein geeignetes KDD-Vorgehensmodell, wird das in Kapitel 3.2 vorgestellte Stufenmodell nach Fayyad et al. (1996) angewendet.

5.1 Fallbeispiel 1: Clusterverfahren im Beschwerdemanagement

In einem Unternehmen findet eine unternehmensübergreifende Produktion von Notebooks statt, d.h. viele verschiedene Einzelteile des Notebooks werden an unterschiedlichen Standorten der Netzwerkpartner gefertigt, geliefert und anschließend schrittweise zusammenmontiert, sodass am Ende der Wertschöpfungskette das finale Produkt entsteht. Aufgrund der steigenden Anzahl von rückgelieferten Produkten und Kundenbeschwerden beschließt das Beschwerdemanagement dieses Problem nicht unbeachtet zu lassen und nach den Ursachen zu suchen. Die Identifikation der Fehlerursachen wird im Unternehmen mit Hilfe des KDD-Modells realisiert. Hierbei wird in den vorliegenden großen Datenbeständen nach Mustern gesucht, die den fehlerhaften Prozessabläufen zugrunde liegen und somit für die Rücksendung der Notebooks verantwortlich sind. Basierend auf dem bekannten Stufenmodell nach Fayyad et al. (1996) beginnt die Datenanalyse mit einer Zieldefinition und einem Domänenverständnis. Dazu verschafft sich der Anwender zunächst einen Überblick über die verschiedenen relevanten Geschäftsabläufe im Unternehmen und setzt sich mit dem Layout der Produktionskette auseinander. Es wird festgehalten, welche Netzwerkpartner welche Produktteile herstellen, weiterverarbeiten oder sie zu Komponenten zusammenmontieren. Hieran anknüpfend wird das Ziel der Datenanalyse definiert, welches darin besteht, Gründe zunächst zu identifizieren, die maßgeblich für den Anstieg der Rücklieferungen sind, und anschließend Maßnahmen zu entwickeln um die Rücklieferungsquote zu verringern. Als nächstes findet im Schritt der Datenselektion eine erste Auseinandersetzung mit dem vorliegenden Datenbestand statt, indem durch Selektion der relevanten Produktdaten von Rücklieferungen der Umfang des Datenbestandes reduziert wird. Hierzu gehören die Daten aller Bauteile, die in jedem rückgelieferten Notebook enthalten sind, wie z.B. Materialnummer, Produktgruppenzugehörigkeit, Auftragsnummer oder Herstellungswerk. Im nächsten Schritt erfolgt die Datenvorbereitung und -bereinigung, die den Großteil des Zeitaufwandes ausmacht. Der selektierte Datenbestand wird nun mit klassischen statistischen Verfahren voruntersucht und behandelt. Bei der Untersuchung fällt auf, dass einige Daten teilweise unvollständig und mehrdeutig sind. Für die Erreichung einer

hohen Datenqualität werden die unvollständigen und mehrdeutigen Daten mit Hilfe geeigneter Verfahren für die nachfolgenden Analyseschritte vorbereitet. Für die Behandlung von Datenlücken werden Mittelwertberechnungen und Interpolationen durchgeführt, wobei für die Behandlung von mehrdeutigen Daten auf Sensitivitätsanalysen zurückgegriffen wird. In einem nächsten Schritt der Datentransformation werden die vorverarbeiteten Daten in eine konsistente Form überführt, indem sie normiert und standardisiert werden. Der Anwender legt besonders viel Wert auf die Qualität der zu analysierenden Daten, denn erfahrungsgemäß sind die Ergebnisse des Modells unzweckmäßig, wenn bereits die Daten selbst nutzlos sind. Als nächstes wird eine geeignete Data Mining-Methode ausgewählt. Dabei ist es von hoher Relevanz die passende Methode mit Blick auf das im Vorfeld definierte Ziel auszuwählen. Um die Einflussfaktoren der defekten Produkte identifizieren zu können, entscheidet sich der Anwender im Rahmen der Segmentierungsmethoden für die Clusteranalyse. Zunächst werden die Ähnlichkeiten und Distanzen der verschiedenen Rücklieferungen bestimmt und quantifiziert. Hierbei werden für alle Rücklieferungen paarweise Vergleiche durchgeführt. Es wird bestimmt, wie ähnlich oder wie unterschiedlich die Merkmale der rückgelieferten Produkte zueinander sind. Anschließend wird aus der Vielzahl existierender Clusterverfahren ein geeigneter Algorithmus ausgewählt. Der Anwender entscheidet sich für das partitionierende Clusterverfahren. Basierend auf den Ähnlichkeiten der Produktmerkmale, werden die einzelnen Datensätze schrittweise in die passende Gruppierung eingeteilt. Es stellt sich heraus, dass sich die Gründe für die Rücklieferungen zum größten Teil in drei verschiedene Cluster aufteilen lassen: (1) Ladebuchse defekt, (2) Grafikkarte defekt und (3) Softwarefehler.

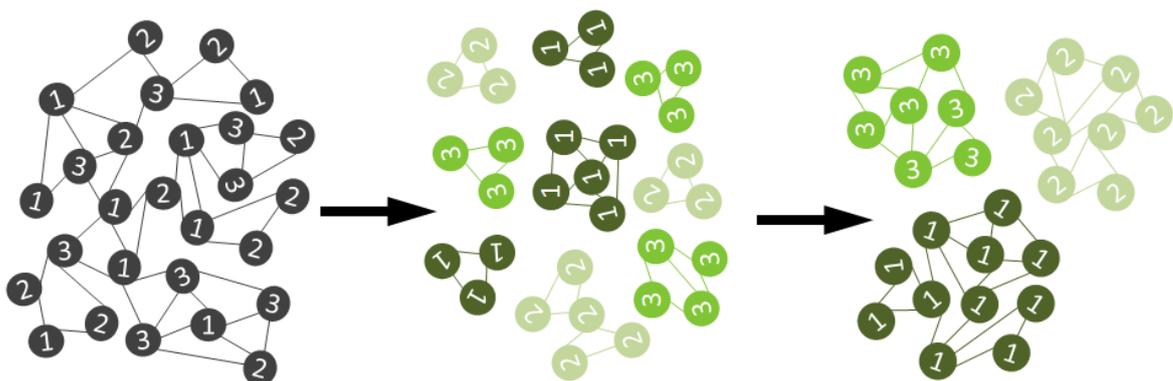


Abb. 11: Cluster der Rücklieferungsgründe

In einem nächsten Schritt wird überprüft, ob die in der Clusteranalyse ermittelten Ergebnisse mit der zuvor festgelegten Zielsetzung vereinbar und in Bezug auf diese interpretierbar sind. Durch die Identifizierung der Hauptgründe für die Rücklieferungen können nun Maßnahmen zur Vermeidung der gängigen Produktfehler entwickelt

werden. Es wird untersucht, welcher Netzwerkpartner für die jeweiligen Bestandteile (Ladebuchse, Grafikkarte und Software) zuständig sind. Abschließend legt das Unternehmen als Konsequenz der Datenanalyse unterschiedliche Maßnahmen fest. Zum einen wird eine erhöhte Kontrolle der Lieferqualität angestrebt, wobei die drei Hauptprodukte Ladebuchse, Grafikkarte und Software nach ihrer Lieferung einer expliziten Funktionsüberprüfung unterzogen werden. Zum anderen wird in Erwägung gezogen, dass bei häufigeren Qualitätsmängeln der angelieferten Produkte ein Wechsel der Netzwerkpartner durchgeführt wird.

5.2 Fallbeispiel 2: Teilebedarfsprognose in der Automobilindustrie

Bei einem Automobilhersteller findet die Produktion von Fahrzeugen durch Just-in-Time-Lieferungen von verschiedenen Bauteilen statt. Dabei werden über den Produktionsprozess hinweg Bauteile bei konkretem Bedarf durch Zulieferer an die jeweiligen Produktionspunkte bzw. -bereiche geliefert, um dort verbaut zu werden. Zur Erzielung von kurzen Lieferzeiten und einer hohen Liefertreue, ist die kontinuierliche Planung und Bereitstellung der zur Produktion benötigten Ressourcen und Bauteile von essentieller Bedeutung. Aufgrund der mit der Lieferzeit und Liefertreue zusammenhängenden Kundenzufriedenheit versucht die Produktionsplanung den Ressourcen- und Bauteilbedarf mit Hilfe des KDD-Modells zu prognostizieren. Basierend auf dem KDD-Stufenmodell nach Fayyad et al. (1996) beginnt der Automobilhersteller mit einem Domänenverständnis sowie der Definition eines Ziels. In diesem Zusammenhang verschafft sich das Unternehmen eingangs einen Überblick über die Produktionsabläufe und die verschiedenen Produktionsbereiche, um so den Verbrauch einzelner Bauteile und deren Kombinationsmöglichkeiten zu identifizieren. Hieran anknüpfend wird das Ziel der Datenanalyse definiert: relative sowie absolute Häufigkeiten zu ermitteln, die den Bedarf an verschiedenen Bauteilen und den dazugehörigen Baugruppen beschreiben. Aus dem zur Verfügung stehenden großen Datenbestand werden die für die Prognose benötigten Daten zunächst selektiert, bereinigt und schließlich transformiert, sodass nach der Datentransformation beispielsweise Daten hinsichtlich der geplanten Fahrzeugstückzahl, der konstruktiv festgelegten Menge einzubauender Komponenten bzw. Teile sowie der Beziehung zwischen bestimmten Merkmalsausprägungen und Bauteilen vorliegen. Die Realisierung der Bedarfsprognose wird im Rahmen der Prognosemethoden des Data Mining mit Neuronalen Netzen vollzogen. Hierbei werden zum einen anhand von funktionalen Zusammenhängen die Abhängigkeiten zwischen den Merkmalsausprägungen des Fahrzeugs (z.B. Ausstattungsmerkmal: Getriebe, Merkmalswert: Automatikgetriebe) und deren Kombinationsmöglichkeiten berücksichtigt. Zum anderen

werden sowohl die geplanten Einbauhäufigkeiten für die Einzelmerkmale als auch das Produktionsprogramm als bekannt vorausgesetzt. Für die Durchführung der Datenanalyse wählt der Anwender einen speziellen Netztypen, welcher mehrere Schichten beinhaltet. Dabei stellen das Produktionsprogramm (z.B. 10 Fahrzeuge von Typ A) sowie die jeweiligen Einzelmerkmale des Fahrzeugs (z.B. automatische Türverriegelung, Klimaanlage) die Eingabeschicht des Neuronalen Netzes dar. Der aktuell berechnete Prognosewert, der den Teilebedarf repräsentiert, wird durch die Ausgabeschicht abgebildet. In den mittleren verdeckten Schichten werden sowohl die Ausgabewerte der vorgelagerten Neuronen zusammengefasst als auch Transformationen durchgeführt, die nichtlinear sind. Somit wird ermöglicht, dass die Zusammenhänge der Ein- und Ausgabewerte durch das Neuronale Netz abgebildet werden können.

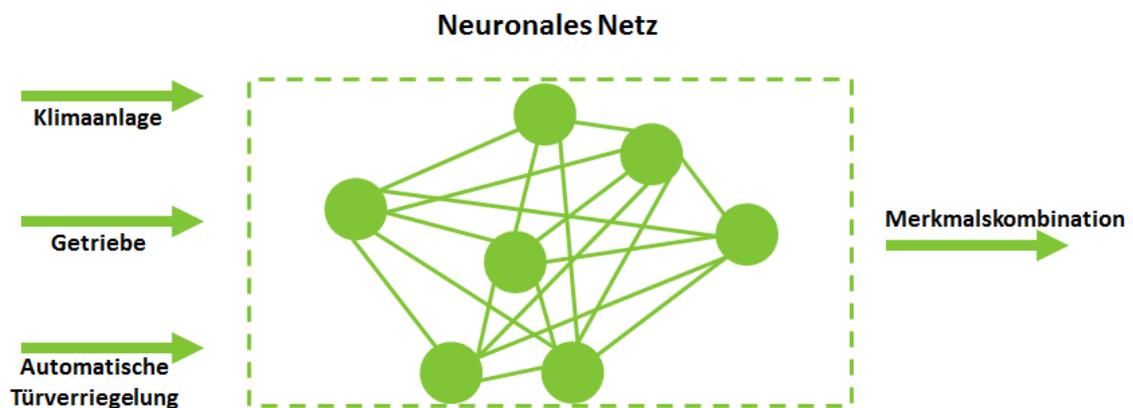


Abb. 12: Neuronales Netz für den Teilebedarf

Der Lernprozess des Netzwerks wird durch Vergangenheitswerte realisiert, wobei eine Approximation des funktionalen Zusammenhangs zwischen der Ein- und Ausgabeschicht durch die Anpassung der Gewichte erfolgt. Nach diesem Data Mining Schritt werden die Analyseergebnisse mit der festgelegten Zieldefinition verglichen. Auffällig ist, dass sich über die eigentliche Zieldefinition hinaus ein weiteres mögliches Ziel ableiten lässt, nämlich die Berechnung der wahrscheinlichsten Kombinationen von Merkmalsausprägungen der Fahrzeuge. Hierzu müssen die einzelnen Analyseschritte wiederholt und im Anschluss analysiert werden. Basierend auf den ermittelten Prognosewerten, wird der Teilebedarf vor der eigentlichen Nutzung im Produktionsprozess an die unterschiedlichen Zulieferer weitergeleitet, sodass eine Just-in-Time-Lieferung gewährleistet wird.

6 Zusammenfassung und Ausblick

Gegenstand der vorliegenden Analyse ist es, eine Zuordnung von problemspezifischen Fragestellungen der Supply Chain Management-Aufgaben und gängigen Data Mining-Methoden zu realisieren.

An die Darstellung relevanter Grundlagen des SCMs, werden basierend auf verschiedene Literaturquellen spezifische Fragestellungen abgeleitet. Als Strukturierungsmöglichkeit dient das weit verbreitete und an verschiedene Supply Chains anpassungsfähige SCOR-Modell. Somit ist es möglich, durch die Prozesselemente Planen, Beschaffen, Herstellen, Liefern und Rückliefern sämtliche Aktivitäten und Fragen, die im Rahmen des Wertschöpfungsprozesses vorkommen, zu kategorisieren. Im weiteren Verlauf der Arbeit werden die Grundlagen des KDD erläutert und als mögliches Beispielmmodell das Stufenmodell nach Fayyad et al. (1996) aufgezeigt. Kennzeichnend für dieses Modell ist die interaktive sowie iterative Prozessstruktur, sodass viele Entscheidungen basierend auf subjektiv bewerteten Informationen getroffen werden. Die angewendeten Data Mining-Methoden dienen demnach als Hilfestellung für den Anwender, der anhand der Zwischenergebnisse den weiteren Ablauf plant. Anschließend wird auf Basis der dargelegten Fragestellungen im SCM und der verschiedenen Wissensentdeckungsmethoden eine tabellarische Gegenüberstellung erarbeitet. Hierzu werden zunächst die Thematik und Problemfelder der einzelnen Fragen dargestellt und jeweils im Anschluss beispielhaft mögliche Zuordnungen von Data Mining-Methoden vorgenommen. Die textuellen Ergebnisse werden schließlich in einer Tabelle im Sinne einer kompakten Übersicht festgehalten. Hierbei wurde die Grundanforderung der Arbeit, das erste Mal eine exemplarische Gegenüberstellung von SCM-Fragestellungen und Data Mining-Methoden auszuarbeiten, in vollem Umfang erfüllt.

Basierend auf den Ergebnissen der Arbeit lässt sich ableiten, dass die praktische Anwendung von Methoden des Data Mining in Zukunft sehr vielfältig eingesetzt werden kann. Dies wird durch die zunehmende Leistungsfähigkeit und wachsende Verbreitung moderner Informationstechnologien ermöglicht. Einer Studie des Fraunhofer IPA (Weskamp et al., 2014) zufolge, bedienen sich bislang hauptsächlich Unternehmen aus Branchen mit starkem direktem Endkundenbezug, wie z.B. Banken, Versicherungen oder Marketing, der Data Mining-Methoden. Trotz der in der dieser Arbeit vorgestellten, möglichen Anwendungsfelder und Ausweitungspotentiale hinsichtlich Produktion und Logistik ist die Verwendung von KDD in dieser Branche weniger verbreitet. Die Herausforderungen liegen zum größten Teil in der Bereitstellung der notwendigen Datenqualität und den damit verbundenen

Datenaufbereitungen. Während 67% der Studienteilnehmer die Sicherstellung der Datenqualität als problematisch bewerten, sehen fast 60% der Teilnehmer in der Bildung von Data Mining Modellen eine große Herausforderung (Weskamp et al., 2014). Allerdings sehen immer mehr Unternehmen und Unternehmensnetzwerke die zunehmende Relevanz der Datenerfassung und -interpretation. Durchgeführte Implementierungen von Data Mining-Methoden in Wertschöpfungsnetzwerke bestätigen, dass die KDD-Ansätze erfolgreich eingesetzt werden und die Unternehmensnetze bei der Gewinnung von Informationen unterstützen können (vgl. Chen et al., 2005; Deuse et al., 2013; Große-Böckmann et al., 2013; Lei/Zhiyong, 2014).

Derzeit ruft das Thema „Industrie 4.0“ in der Industrie hohe Aufmerksamkeit und hohes Interesse hervor. Hierbei steht die Betrachtung einer wachsenden informationstechnischen Vernetzung innerhalb von Wertschöpfungsprozessen im Vordergrund. Gerade in diesem Umfeld kann die vorliegende Arbeit einen guten Baustein bilden, denn die zunehmende Integration der Informationstechnik in die Produktion und Logistik führt zur Erzeugung großer Datenmengen, die nach dem vorgestellten KDD-Ansatz bearbeitet und analysiert werden können. Die automatische Erfassung von Daten wird durch hochmoderne und sensorreiche Produktionsanlagen und Produkte realisiert, wobei bei der Datenerfassung eine hohe Datenqualität angestrebt wird. Somit ist ersichtlich, dass die Entwicklungstendenzen hinsichtlich Industrie 4.0 die Relevanz sowie die Ausweitung der KDD-Modelle in Wertschöpfungsnetzwerken beeinflussen und ebenso steigern werden (Schöning/Dorchain, 2014). Durch die hohen Potenziale und neuen Anwendungsfelder der Data Mining-Methoden können sich im Rahmen der Supply Chains gute Möglichkeiten ergeben, die zu einer datengestützten Optimierung von Materialflüssen und Produktionsabläufen beitragen können.

Literaturverzeichnis

- Agrawal, R., Imielinski, T., Swami, A. (1993): Mining Association Rules between Sets of Items in Large Databases, in: SIGMOD '93 - Proceedings of the 1993 ACM SIGMOD international conference on Management of data, S. 207-216.
- Arndt, H. (2010): Supply Chain Management - Optimierung logistischer Prozesse, 5. Aufl., Gabler Verlag, Wiesbaden.
- Backhaus, K., Erichson, B., Plinke, W., Weiber, R. (2008): Multivariate Analysemethoden - Eine anwendungsorientierte Einführung, 12. Aufl., Springer Verlag, Berlin Heidelberg.
- Bankhofer, U. (2004): Data Mining und seine betriebswirtschaftliche Relevanz, in: Betriebswirtschaftliche Forschung und Praxis, Vol. 56, Heft 4, S. 395 - 412.
- Baumgarten, H. (2004): Entwicklungsphasen des Supply Chain Managements, in: Baumgarten, H., Darkow, I.-L., Zadek, H. (Hrsg.): Supply Chain Steuerung und Services, Springer Verlag, Berlin Heidelberg, S. 51-60.
- Becker, J., Beverungen, D., Knackstedt, R. (2008): Wertschöpfungsnetzwerke von Produzenten und Dienstleistern als Option zur Organisation der Erstellung hybrider Leistungsbündel, in: Becker, J., Knackstedt, R., Pfeiffer, D. (Hrsg.): Wertschöpfungsnetzwerke - Konzepte für das Netzwerkmanagement und potenzielle aktueller Informationstechnologien, Physica Verlag, Heidelberg, S. 3 - 31.
- Becker, M. (2010): Analyse der Einsatzmöglichkeiten von Business-Intelligence-Systemen zur strategischen Entscheidungsunterstützung in Wertschöpfungsnetzwerken, Diplomica Verlag, Hamburg.
- Beckmann, H. (2012): Prozessorientiertes Supply Chain Engineering - Strategien, Konzepte und Methoden zur modellbasierten Gestaltung, Springer Verlag, Wiesbaden.
- Betge, D. (2006): Koordination in Advanced Planning und Scheduling-Systemen, Deutscher Universitätsverlag, Wiesbaden.
- Bolstorff, P. A., Rosenbaum, R. G., Poluha, R. G. (2007): Spitzenleistungen im Supply Chain Management - Ein Praxishandbuch zur Optimierung SCOR, Springer Verlag, Berlin Heidelberg.
- Chamoni, P., Beekmann, F., Bley, T. (2013): Ausgewählte Verfahren des Data Mining, in: Chamoni, P., Gluchowski, P. (Hrsg.): Analytische Informationssysteme - Business Intelligence-Technologien und -Anwendungen, 4. Aufl., Springer Verlag, Berlin Heidelberg, S. 329-358.

- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. (2000): CRISP-DM 1.0 - Step-by-Step data mining guide, in: URL: <ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf> [aufgerufen am 05.05.2015].
- Chen, M.-C., Huang, C.-L., Chen, K.-Y., Wu, H.-P. (2005): Aggregation of orders in distribution centers using data mining, in: Expert Systems with Applications, Vol. 28, Heft 3, S. 453 - 460.
- Cleve, J., Lämmel, U. (2014): Data Mining, Oldenbourg Verlag, München.
- Corsten, H., Gössinger, R. (2008): Einführung in das Supply Chain Management, 2. Aufl., Oldenbourg Verlag, München.
- Deuse, J., Erohin, O., Lieber, D. (2013): Wissensentdeckung im industriellen Kontext - Herausforderungen und Anwendungsbeispiele, in: Zeitschrift für wirtschaftlichen Fabrikbetrieb, Vol. 108, Heft 6, S. 388 - 393.
- Düsing, R. (2000): Knowledge Discovery in Databases, in: Wirtschaftsinformatik, Vol. 42, Heft 1, S. 74 – 75.
- Ertel, W. (2013): Grundkurs Künstliche Intelligenz - Eine praxisorientierte Einführung, 3. Aufl., Springer Vieweg Verlag, Wiesbaden.
- Eßig, M., Hofmann, E., Stölzle, W. (2013): Supply Chain Management, Franz Vahlen Verlag, München.
- Fandel, G., Stütz, S., Fistek, A. (2011): Produktionsmanagement, 2. Aufl., Springer Verlag, Berlin Heidelberg.
- Fayyad, U. M.; Piatetsky-Shapiro, G.; Smyth, P. (1996): From Data Mining to Knowledge Discovery in Databases, in: Communications of the ACM, Vol. 39, Heft. 11, S. 37 - 54.
- Fisher, M. L. (1997): What is the right Supply Chain for your Product?, in: Harvard Business Review, Vol. 75, Heft 2, S. 105 - 116.
- Fockel, R. (2009): Methoden des Data Mining im praktischen Einsatz, Shaker Verlag, Aachen.
- Gabriel, R., Gluchowski, P., Pastwa, A. (2009): Data Warehouse und Data Mining, W3L-Verlag, Witten Herdecke.
- Gluchowski, P., Gabriel, R., Dittmar, C. (2008): Management Support Systeme und Business Intelligence - Computergestützte Informationssysteme für Fach- und Führungskräfte, 2. Aufl., Springer Verlag, Berlin Heidelberg.

- Große-Böckmann, M., Krappig, R., Stolorz, M., Schmitt, R. (2013): Data-Mining in der Produktion - Neue Methoden für eine robuste Prozessentwicklung, in: Werkstattstechnik online, Vol. 103, Heft 11/12, S. 921 - 925.
- Grünauer, K. M.; Fleisch, E.; Österle, H. (2000): Information Systems for Supply Chain Management – An Overview, in: Fleisch, E.; Österle, H., Alt, R. (Hrsg.): Business Networking - Shaping Enterprise Relationships on the Internet, Springer Verlag, Berlin Heidelberg, S. 185 - 200.
- Harrington, L. (1995): Logistics, Agent for Change - Shaping the Integrated Supply Chain, in: Journal of Transportation & Distribution, Vol. 36, Heft 1, S. 30-34.
- Hellingrath, B., Hegmanns, T., Maaß, J.-C., Toth, M. (2008): Prozesse in Logistiknetzwerken - Supply Chain Management, in: Arnold, D., Isermann, H., Kuhn, A., Tempelmeier, H., Furmans, K. (Hrsg.): Handbuch Logistik, 3. Aufl., Springer Verlag, Berlin Heidelberg, S. 459 - 486.
- Hieronimus, M. (2005): Strategisches Controlling von Supply Chains – Entwicklung eines ganzheitlichen Ansatzes unter Einbeziehung der Wertschöpfungspartner, Cuvillier Verlag, Göttingen.
- Hilpert, K. (2014): Einsatz maschineller Lernverfahren im Decision Support von Wertschöpfungsnetzwerken, Dortmund.
- Kiener, S., Maier-Scheubeck, N., Obermaier, R., Weiß, M. (2012): Produktions-Management - Grundlagen der Produktionsplanung und -steuerung, 10. Aufl., Oldenbourg Verlag, München.
- Kießwetter, M., Vahlkamp, D. (2007): Data Mining in SAP NetWeaver BI, Galileo Press Verlag, Bonn.
- Kuhn, A., Hellingrath, B. (2002): Supply Chain Management - Optimierte Zusammenarbeit in der Wertschöpfungskette, Springer Verlag, Berlin Heidelberg.
- Kurbel, K. (2011): Enterprise Resource Planning und Supply Chain Management in der Industrie, 7. Aufl., Oldenbourg Verlag, München.
- Lawrenz, O., Hildebrand, K., Nenninger, M. (2000): Supply Chain Management. Strategien, Konzepte und Erfahrungen auf dem Weg zu E-Business Networks, Springer Verlag, Berlin Heidelberg.
- Lei, C., Zhiyong, T. (2014): Research of RFID Data Mining Based on Supply Chain Management, in: IEEE Workshop on Electronics, Computer and Applications, S. 668 - 671.

- Meyr, H., Rosic, H., Seipl, C., Wagner, M., Wetterauer, U. (2008): Architecture of Selected APS, in: Stadtler, H., Kilger, C. (Hrsg.): Supply Chain Management and Advanced Planning - Concepts, Models, Software and Case Studies, 4. Aufl., Springer Verlag, Berlin Heidelberg, S. 349 - 366.
- Nakhaeizadeh, G. (1999): Data Mining - Theoretische Aspekte und Anwendungen, Physica Verlag, Heidelberg.
- Porter, M. E. (2006): Creating Tomorrow`s Advantages, in: Hahn, D., Taylor, B. (Hrsg.): Strategische Unternehmensplanung - Strategische Unternehmensführung - Stand und Entwicklungstendenzen, 9. Aufl., Springer Verlag, Berlin Heidelberg, S. 267-274.
- Porter, M. E. (2010): Wettbewerbsvorteile - Spitzenleistungen erreichen und behaupten, 7. Aufl., Campus Verlag, Frankfurt New York.
- Schöning, H., Dorchain M. (2014): Data Mining und Analyse, in: Bauernhansl, T., ten Hompel, M., Vogel-Heuser, B. (Hrsg.): Industrie 4.0 in Produktion, Automatisierung und Logistik, Springer Vieweg Verlag, Wiesbaden, S. 543 - 554.
- Supply Chain Council (2011): SCOR Framework, in: URL: <http://www.apics.org/sites/apics-supply-chain-council/frameworks/scor> [aufgerufen am 20.05.2015].
- Towill, D. R. (1996): Time compression and supply chain management - a guided tour, in: Logistics Information Management, Vol. 9, Heft 6, S. 41 - 53.
- Van Eck, M. (2003): Advanced Planning and Scheduling - Is logistics everything?, BWI Paper, Amsterdam.
- Werner, H. (2013): Supply Chain Management - Grundlagen, Strategien, Instrumente und Controlling, 5. Aufl., Gabler Verlag, Wiesbaden.
- Weskamp, M., Tamas, A., Wochinger, T., Schatz, A. (2014): Einsatz und Nutzenpotenziale von Data Mining in Produktionsunternehmen, Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA, Stuttgart.
- Wiedmann, K.-P., Buckler, F. (2001): Neuronale Netze im Management, in: Wiedmann, K.-P., Buckler, F. (Hrsg.): Neuronale Netze im Marketing-Management - Praxisorientierte Einführung in modernes Data-Mining, Gabler Verlag, Wiesbaden, S. 35 - 100.
- Wiedmann, K.-P., Buckler, F., Buxel, H. (2001): Data-Mining - ein einführender Überblick, in: Wiedmann, K.-P., Buckler, F. (Hrsg.): Neuronale Netze im Marketing-Management - Praxisorientierte Einführung in modernes Data-Mining, Gabler Verlag, Wiesbaden, S. 15-33.

Anhang

Eidesstaatliche Erklärung

Gürez, Ebru

Name, Vorname

141886

Matr.-Nr.

Ich versichere hiermit an Eides statt, dass ich die vorliegende Bachelorarbeit/Masterarbeit* mit dem Titel

Zuordnung von Data Mining-Methoden zu problemspezifischen Fragestellungen
von Supply Chain Management-Aufgaben

selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Dortmund, den 10. Juni 2015

Ort, Datum

Unterschrift

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -)

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird gfls. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Dortmund, den 10. Juni 2015

Ort, Datum

Unterschrift