

# Data Mining im Retourenmanagement: Evaluation von Retourenmengenprognosen anhand der Transaktionsdaten eines Schuh- und Bekleidungsversandhändlers

David Karl

Lehrstuhl für Betriebswirtschaftslehre, insb. Produktion und Logistik, Otto-Friedrich-Universität Bamberg / Kompetenzzentrum für Geschäftsmodelle in der digitalen Welt, Feldkirchenstr. 21, 96052 Bamberg, david.karl@uni-bamberg.de

1	Einleitung.....	191
2	Begriffliche und methodische Grundlagen.....	192
3	Data Mining-Modelle zur Prognose der Retourenentscheidung .....	199
4	Evaluation von Retourenmengenprognosen.....	204
5	Fazit .....	210
6	Literaturverzeichnis .....	211

## Abstract:

*Durch die Digitalisierung aller Lebensbereiche gewinnt der Online-Handel an Bedeutung: Steigende Versandhandelsumsätze und die zunehmende Unabhängigkeit von einem stationären Point of Sale führen zu höheren Logistikaufkommen und folglich zu einer größeren Gütermobilität. Dies gilt aufgrund gesetzlicher Vorgaben in Verbindung mit liberalen Rücknahmeregelungen insbesondere für die Rückwärtslogistik und begründet operative Herausforderungen im Bereich des Bestands- und Retourenmanagements. Besonders von der Retourenproblematik betroffen ist die Fashion-Branche, in der Kunden nicht selten jedes zweite Paket zurücksenden.*

*Dieser Beitrag leitet zunächst aus umfangreichen historischen Transaktionsdaten eines deutschen Schuh- und Bekleidungsversandhändlers mit Hilfe von Data Mining-Methoden Modelle ab, die zur Prognose zukünftiger Retouren schon zum Bestellzeitpunkt dienen. Die Evaluation der mit diesen Modellen berechneten mengenbezogenen Vorhersage im Vergleich zu gängigen Prognoseverfahren erlaubt Rückschlüsse auf den Nutzen von Data Mining im Retourenmanagement. Zudem lassen sich praxisnahe Empfehlungen generieren, welche Methoden für die Retourenmengenprognose am geeignetsten sind.*

## 1 Einleitung

„Einzelhandelsunternehmen [...] [haben] längst begriffen, dass die Digitalisierung all unserer Geschäftsprozesse das jahrhundertalte Geschäftsmodell des Handels komplett auf den Kopf stellen wird. Besser gesagt, die Digitalisierung wird aus dem Handel in den nächsten Jahren ein Experimentierfeld für entfesselte Datenanalysen aller Art machen.“<sup>1</sup>

Dieses Zitat illustriert, dass die Datenanalyse einen kritischen Erfolgsfaktor für den Handel darstellt und darstellen wird. Nachfrageprognosen – oft ohne informationstechnische Unterstützung intuitiv geschätzt<sup>2</sup> – waren für den stationären Handel schon lange essenziell, um Kundenbedürfnisse abschätzen und befriedigen zu können. Die aktuell stattfindende digitale Transformation von Prozessen und ganzen Geschäftsmodellen durch Nutzung moderner Informations- und Kommunikationstechnologien erweitert für Händler die potentielle Datenbasis und die Analysemöglichkeiten: Betriebliche Informationssysteme wie Warenwirtschafts- oder ERP-Systeme halten große Datensätze vor, die bei der Abwicklung von Geschäftsprozessen entweder als „Nebenprodukt“ anfallen oder bewusst erhoben werden (z. B. geäußerte Kundenpräferenzen). Insbesondere die Formulierung „Experimentierfeld“<sup>3</sup> aus dem eingangs erwähnten Zitat deutet jedoch auf große Unklarheiten hin, wie Händler diese Daten sinnvoll und nutzenbringend aus- und verwerten können.

Vor allem der Versandhandel, dessen Geschäftsmodell aufgrund des räumlichen Auseinanderfallens von Angebot und Nachfrage in besonderer Weise durch Informationsflüsse zum und vom Kunden<sup>4</sup> geprägt ist, kann von einer gezielten Datennutzung profitieren. Gegenüber dem stationären Handel gewinnt er kontinuierlich an Bedeutung.<sup>5</sup> In der Schuh- und Bekleidungsbranche entfallen bereits knapp 20 % des Umsatzes auf den Online-Handel.<sup>6</sup> Da das Inspizieren des interessierenden Artikels im Laden entfällt, haben Kunden gemäß EU-Verbraucherrechterichtlinie 2011/83/EU das Recht, bestellte Waren zu Hause zu begutachten und wieder zurückzusenden.<sup>7</sup> Aufgrund der damit verbundenen Aufwendungen und möglichen Umsatzausfälle für den Händler ist die Planung, Steuerung und Kontrolle dieser Retouren „[...] elementarer Bestandteil des [Distanzhandels-]Geschäftsmodells.“<sup>8</sup> We-

---

<sup>1</sup> Dziemba/Wenzel (2014), S. 95.

<sup>2</sup> Vgl. Crone (2010), S. 92.

<sup>3</sup> Dziemba/Wenzel (2014), S. 95.

<sup>4</sup> Wenn im Folgenden aus Gründen der besseren Lesbarkeit nur von Kunden, Käufern etc. die Rede ist und keine explizite Definition des Geschlechts erfolgt, sind stets sowohl männliche als auch weibliche Gruppenmitglieder gemeint.

<sup>5</sup> Vgl. Jahn (2013), S. 3; Handelsverband Deutschland (2014a).

<sup>6</sup> Vgl. Handelsverband Deutschland (2014b).

<sup>7</sup> Vgl. Europäische Union (2011), S. 6.

<sup>8</sup> Asdecker (2014), Klappentext hinten.

gen einer zunehmenden Verschmelzung von stationärem und Online-Handel können sich klassische Einzelhändler vor dieser Problematik nicht verschließen und müssen sich Gedanken über Rücknahme- und Rücksendepolitiken machen.<sup>9</sup>

Eine Befragung von im Retourenmanagement tätigen Praktikern durch Asdecker (2014) bezüglich vorliegender Forschungslücken und der damit verbundenen Praxisbedeutung stützt die Relevanz der Datenanalyse im Retourenmanagement: Viele Entscheidungsträger sehen im Rahmen der taktischen Planung, einer maßgeblichen Komponente des Retourenmanagements<sup>10</sup>, Forschungsbedarf zur „Abschätzung [der] Retourenmenge“<sup>11</sup> durch „Prognosemodelle“<sup>12</sup>.

Auf Basis der realen Transaktionsdaten eines Händlers aus der Schuh- und Bekleidungsbranche über den Zeitraum eines Jahres sollen die folgenden Fragestellungen beantwortet werden:

- Wie präzise können Händler auf Basis von Transaktionsdaten die Rücksendeentscheidung eines Kunden mit Hilfe von Data Mining-Methoden prognostizieren?
- Welche Verfahren eignen sich am besten zur Prognose von Retourenmengen?
- Ermöglichen die Data Mining-Modelle zur Retourenentscheidung die Ableitung einer präzisen Retourenmengenprognose?

Hierzu führt Kapitel 2 kurz in die begrifflichen und methodischen Grundlagen des Retourenmanagements und des Data Mining ein. Kapitel 3 stellt den verwendeten Datensatz des Versandhändlers und darauf basierende Data Mining-Modelle zur Prognose der Rücksendeentscheidung vor. Kapitel 4 nutzt diese Modelle, um daraus Vorhersagen zur Menge der rückgesendeten Artikel und Pakete zu erstellen, und vergleicht diese mit naiven Schätzungen der Rücksendemenge auf Basis von Absatzzahlen und historischen Retourenquoten. Das abschließende Kapitel 5 fasst die dabei erzielten Ergebnisse kurz zusammen, zeigt die Limitationen der Untersuchung auf und benennt zukünftigen Forschungsbedarf.

## 2 Begriffliche und methodische Grundlagen

Abschnitt 2.1 führt zunächst grundlegend in das Themengebiet des Retourenmanagements ein, erläutert die für den weiteren Fortgang relevanten Kennzahlen und zeigt unter 2.2 auf, wo die Nutzenpotentiale der Datenanalyse in diesem Bereich liegen. Die begrifflichen Grundlagen des Data Mining sowie für diesen Beitrag einschlägige Verfahren skizziert Abschnitt 2.3, während Abschnitt 2.4 allgemein gängige einfache Prognoseverfahren kurz zusammenfasst. Unter 2.5 findet sich eine

---

<sup>9</sup> Vgl. z. B. Dziemba/Wenzel (2014), S. 112; Jahn (2013), S. 8; ebd., S. 13.

<sup>10</sup> Vgl. Asdecker (2014), S. 144.

<sup>11</sup> Asdecker (2014), S. 160.

<sup>12</sup> Asdecker (2014), S. 160.

Übersicht über verbreitete Maßzahlen zur Evaluation verschiedener Prognosemodelle, die im weiteren Verlauf der Abhandlung zum Einsatz kommen.

### 2.1 Retourenmanagement und dafür relevante Kennzahlen

Der Begriff der Retoure umfasst grundsätzlich alle Rücksendungen an vorgelagerte Institutionen des Wertschöpfungsprozesses oder beauftragte Dienstleister.<sup>13</sup> Das Retourenmanagement bezeichnet die „[...] institutionsübergreifende Planung, Steuerung und Kontrolle der Retourenflüsse sowie der damit assoziierten Informations- und Finanzflüsse, mit dem Ziel die Gewinnmaximierung des Wertschöpfungssystems zu unterstützen.“<sup>14</sup>

Die Aufgaben des Retourenmanagements unterteilen sich in zwei Bereiche: einerseits auf taktischer Ebene **präventive** Aufgaben (Verhinderung, Vermeidung, Förderung), die eine frühzeitige Beeinflussung vor dem Entstehen einer Rücksendung beinhalten, andererseits überwiegend auf operativer Ebene **kurative** Aufgaben zur optimalen Abwicklung des Retourenprozesses, welche faktisch vorhandene oder nicht zu verhindernde Rücksendungen fokussieren.<sup>15</sup> Dieser Beitrag betrachtet schwerpunktmäßig das kurative, unternehmensinterne Retourenmanagement und beschränkt sich zudem auf Rücksendungen vom Endkunden zum Händler, also C2B-Transaktionen („consumer product returns“).

Kennzahlen dienen im Allgemeinen als verdichtete betriebliche Informationen dazu, Managemententscheidungen mit einer Datengrundlage zu untermauern und die Komplexität bei der Interpretation der geschäftlichen Realität zu verringern.<sup>16</sup> Auch Rücksendungen lassen sich erst mithilfe von aggregierten Daten sinnvoll planen, steuern und kontrollieren, wie diese häufig genannte Managerweisheit verdeutlicht: „[I]f you can't measure it, you can't manage it.“<sup>17</sup>

Die wohl wichtigste Kennzahl für das Retourenmanagement<sup>18</sup> errechnet sich aus dem Verhältnis zurückgesendeter Einheiten zur Summe versendeter Einheiten:

$$\text{Retourenquote (RQ)} = \frac{\text{zurückgesendete Einheiten}}{\text{versendete Einheiten}}$$

Sie lässt sich insbesondere für die Zwecke des Distanzhandels in  $\alpha$ -,  $\beta$ - und  $\gamma$ -Retourenquote differenzieren, wobei jede Sicht bestimmte unternehmerische Funktionsbereiche hervorhebt.<sup>19</sup> Die  **$\alpha$ -Retourenquote** ( $\alpha$ -RQ) bezeichnet das Verhältnis

<sup>13</sup> Vgl. Asdecker (2011), S. 258; Rogers et al. (2002), S. 3f.

<sup>14</sup> Asdecker (2012), S. 495.

<sup>15</sup> Vgl. Rogers et al. (2002), S. 5; Asdecker (2014), S. 23.

<sup>16</sup> Vgl. Weber/Wallenburg (2010), S. 332f.

<sup>17</sup> Garvin (1993), S. 89. Auch an anderen Stellen mit identischem Wortlaut und ohne Quellennachweis zitiert, z. B. Kaplan/Norton (1996), S. 21.

<sup>18</sup> Vgl. Asdecker (2014), S. 229f.

<sup>19</sup> Vgl. für den folgenden Absatz Asdecker (2013), S. 42–44.

von zurückgesendeten zu versendeten **Paketen** und legt den Fokus somit auf die logistische Perspektive, während die  **$\beta$ -Retourenquote** ( $\beta$ -RQ) dieses Verhältnis in Bezug auf einzelne **Artikel** ausdrückt und die Sichtweise des Marketings einnimmt. Die  **$\gamma$ -Retourenquote** ( $\gamma$ -RQ) beschreibt das monetäre, **wertmäßige** Verhältnis zurückgesendeter zu versendeten Einheiten, was der Finanz- und Controlling-Perspektive entspricht. Die weitere Analyse betrachtet aufgrund der operativen Schwerpunktsetzung dieses Beitrags die  $\alpha$ - und  $\beta$ -Retourenquoten.

Eine detailliertere Einteilung der Retourenquote lässt sich in Bezug auf die Sichtweise oder Aggregationsebene durchführen. Gängige Bezugseinheiten sind Kunden, Artikel, Artikelvarianten (z. B. Farben), Hersteller oder eine Gesamtbetrachtung aus Unternehmenssicht. Zur Vermeidung von Verwechslungen ist diese Referenz stets mit anzugeben, was zu zweidimensionalen Ausprägungen dieser Kennzahl führt.

Neben der Retourenquote nutzen Praxisvertreter weitere Kennzahlen wie z. B. die absolute Anzahl der Retouren, Durchlaufzeiten, Retourengründe oder Fehlerquoten.<sup>20</sup> Diese Untersuchung beschäftigt sich mit aus den historischen Transaktionsdaten abgeleiteten operativen Verkaufs-Kennzahlen und untersucht deren Einfluss für das Retourenmanagement.

## 2.2 Kontextbezogene Nutzenpotentiale der Datenanalyse

Datenanalysen zum Managementzweck intendieren einen betriebswirtschaftlichen Nutzen, also im weiteren Sinne Wertschöpfung. Der folgende Abschnitt skizziert Verwertungspotentiale, die sich auf Basis der retourenbezogenen Datenauswertung im Versandhandel heben lassen.

Bei der Vorhersage einer Rücksendung interessiert die Tatsache, ob ein Kunde einen Artikel bzw. ein Paket zurücksendet, der Zeitpunkt, wann er diese Rücksendung tätigt, und der Zustand der zurückgesendeten Artikel. Der *Rücksendezeitpunkt* betrifft insbesondere die Ressourcendisposition, aber auch die Bestellmengenplanung. Der *Artikelzustand* beeinflusst die Wiederverkäuflichkeit des Artikels und den Retourenbearbeitungsprozess, also die nötigen Verwertungs-, Aufbereitungs- oder Entsorgungsschritte, was sich auf Prozesskosten und Bestellmengen bei den Lieferanten auswirkt. Über Zustand und Zeitpunkt einer Rücksendung lassen sich anhand der vorliegenden Transaktionsdaten keinerlei Aussagen treffen. Die weiteren Ausführungen beziehen sich auf die reine *Retourenentscheidung* und deren Wahrscheinlichkeit.

Die Auswertung der paketbezogenen Rücksendewahrscheinlichkeit bzw. der  $\alpha$ -Retourenquote ist eher im kurativen Retourenmanagement anzusiedeln, da diese Kennzahl und die sich daraus ergebende Anzahl an Paketen die Kapazitätsplanung

---

<sup>20</sup> Vgl. Asdecker (2014), S. 227–231.

für konkrete logistische Operationen unterstützt, beispielsweise in der Personaleinsatzplanung.<sup>21</sup>

Eine Vorhersage der artikelbezogenen Retourenmenge besitzt operativ/kurativ für die Losgrößen- und Bestellmengenplanung Entscheidungsrelevanz, weil sich – unter Annahme der Wiederverwendbarkeit – Rücksendungen direkt auf Bestände auswirken. Deren Management stellt in der Praxis eine der Hauptproblematiken für den Versandhandel dar: „Die zentrale Frage für die Händler ist, wie sie ihre Bestände und die Anzahl der Retouren in den Griff bekommen.“<sup>22</sup> Präzise Prognosen für Rücksendungen können Über- oder Unterbestände verhindern. Erstere haben erhöhte Lagerkosten und eine längere Kapitalbindung zur Folge.<sup>23</sup> Unterbestände führen zu Umsatzausfällen und verursachen Opportunitätskosten, ziehen höhere Beschaffungskosten nach sich und bewirken Verärgerung beim Kunden (Goodwill-Verluste).<sup>24</sup> Artikelspezifische Retoureninformationen können zudem bei Lieferantenverhandlungen von Nutzen sein, um höhere Rabatte, verbesserte Qualitätsstandards oder niedrigere Fertigungstoleranzen durchzusetzen.

### 2.3 Data Mining-Grundlagen

Data Mining kombiniert laut Ratner (2012) drei Bestandteile: klassische statistische Methoden der explorativen Datenanalyse, Big Data sowie maschinelles Lernen mithilfe von PCs.<sup>25</sup> Somit ist keine trennscharfe Abgrenzung des Data Mining vom Big Data-Begriff möglich. Es geht beim „Mining“ mehr um den Prozess, der aus komplexen Datensätzen mittels klassischer statistischer oder moderner informationstechnischer Algorithmen werthaltige Informationen zu gewinnen versucht, wie Brethenoux von der Gartner Group zusammenfasst:

„[Data Mining ist der] Prozess des Entdeckens bedeutsamer neuer Zusammenhänge, Muster und Trends durch die Analyse großer Datensätze mittels Mustererkennung sowie statistischer und mathematischer Verfahren.“<sup>26</sup>

Data Mining-Verfahren lassen sich nach Anwendungszweck überblicksmäßig in vier Kategorien einteilen, wobei die Einordnungen nicht trennscharf sind, da identische Algorithmen in mehreren Bereichen zum Einsatz kommen können: Numerische Ver-

<sup>21</sup> Vgl. Asdecker (2014), S. 261.

<sup>22</sup> Loebich/BearingPoint (2014).

<sup>23</sup> Vgl. für den folgenden Absatz Feindt/Sinn (2014), S. 24.

<sup>24</sup> Vgl. Werner (2013), S. 229; Feindt/Jarke (2014), S. 330.

<sup>25</sup> Vgl. Ratner (2012), S. 13f.

<sup>26</sup> Brethenoux (unbekanntes Jahr), zitiert nach Kuß/Wildner/Kreis (2014), S. 33. Diese Definition von Brethenoux ist weit verbreitet, wird jedoch stets ohne genaue Quellenangabe zitiert, weswegen ein exakter Quellennachweis nicht möglich ist.

fahren, Klassenbildungsverfahren, Assoziationsanalysen und Klassifikationsverfahren, von denen nur letztere für den weiteren Gang der Untersuchung relevant sind.<sup>27</sup>

Solche Klassifikationsverfahren, zu denen u. a. Diskriminanzanalysen, Klassifikations- und Entscheidungsbäume oder neuronale Netze gehören, bestimmen aufgrund einer Menge von Prädiktoren die Klassenzugehörigkeit für eine nominalskalierte (im Extremfall dichotome) abhängige Variable.<sup>28</sup> Neben der Klassenzugehörigkeit ist deren Wahrscheinlichkeit oder die Vorhersagekonfidenz von Interesse.<sup>29</sup> Im Gegensatz zu Clusteranalysen sind die Klassen a priori bekannt, teils bedienen sich beide Verfahren aber derselben Algorithmen, z. B. der bereits genannten neuronalen Netze.

Die Kombination verschiedener Data Mining-Verfahren zu hybriden Modellen kann Schwächen eines einzelnen Modells ausgleichen und die Effizienz des Gesamtmodells verbessern, zumal Data Mining-Softwarepakete die automatisierte Modellverknüpfung ermöglichen.<sup>30</sup> Eine Kreuzvalidierung des Modells durch Ziehen und Evaluieren verschiedener Teilstichproben und Aggregation mehrerer gelernter Teilmodelle kann die Vorhersage weiter stabilisieren oder verbessern.<sup>31</sup>

Für die vorliegende Untersuchung kommen die Softwarepakete IBM SPSS Statistics 23 (klassisches Statistik-Paket), IBM SPSS Modeler 17 (Data Mining-Tool) sowie Microsoft Excel 2013 (alle numerischen Berechnungen) zum Einsatz.

#### 2.4 Gängige Prognoseverfahren

Prognosen oder Vorhersagen bezeichnen Aussagen über zukünftige Ereignisse und Entwicklungen auf qualitativer oder quantitativer Grundlage, die rationale Entscheidungen fördern.<sup>32</sup> Verfahren der qualitativen Prognose, welche eher auf die Abschätzung langfristiger Entwicklungen zielt, sind bspw. Expertenbefragungen (u. a. Delphi-Methode) oder Szenariotechniken, die für die vorliegende Thematik jedoch nicht weiter relevant sind.<sup>33</sup> Die quantitative, kurz- bis mittelfristig orientierte Prognoserechnung lässt sich in zwei Teilbereiche systematisieren: Die Zeitreihenextrapolation einerseits bildet Zukunftswerte nur aus den Vergangenheitswerten derselben Variable und berücksichtigt keine Einflüsse außer der Zeit, was gleichzeitig auch großer Kritikpunkt solcher Verfahren ist.<sup>34</sup> Kausalanalysen andererseits leiten die inte-

<sup>27</sup> Vgl. Petersohn (2005), S. 25; Cleve/Lämmel (2014), S. 57.

<sup>28</sup> Vgl. Küsters (2001), S. 107.

<sup>29</sup> Vgl. Cleve/Lämmel (2014), S. 61.

<sup>30</sup> Vgl. Fidan/Ozkok (2013), S. 121. Trotz des dortigen Einsatzes im Rahmen von Aktienprognosen lässt sich diese Aussage auf den Kontext des Data Minings im Retourenmanagement übertragen.

<sup>31</sup> Vgl. Domingos (2012), S. 85f; Fromm (2005), S. 23.

<sup>32</sup> Vgl. Hansmann (1983), S. 7; ebd., S. 12; Rosentreter (1977), S. 5.

<sup>33</sup> Vgl. Hansmann (1983), S. 18–26.

<sup>34</sup> Vgl. Hansmann (1983), S. 45.

ressierende Größe anhand von erklärenden Variablen her, berücksichtigen dabei Zusammenhänge und ermöglichen auch die Deutung der Wirkungsrichtungen und -stärken.

Einfache **Zeitreihenprognosen** wie ein gleitender Mittelwert oder die einfache exponentielle Glättung bilden die erste untersuchte Verfahrensklasse. Diese beiden Verfahren gehen von einem konstanten Modell aus, d. h. der zu prognostizierende Wert unterliegt annahmegemäß keinem grundlegenden Trend oder einer Saisonalität. Der in Periode  $T$  erstellte Prognosewert  $\hat{y}_{T,T+1}$  für Periode  $T + 1$  auf Basis des gleitenden Mittelwerts (GMW) berechnet sich aus dem arithmetischen Mittelwert der letzten  $N$  Perioden:<sup>35</sup>

$$\hat{y}_{T,T+1} = \frac{1}{N} \cdot \sum_{t=T+1-N}^T y_t$$

Das Prinzip der einfachen exponentielle Glättung (1EXPO) zieht *alle* Vergangenheitswerte für die Prognose heran, allerdings mit exponentiell sinkender Gewichtung: Je länger ein Wert in der Vergangenheit liegt, desto geringer beeinflusst er den nächsten Prognosewert.<sup>36</sup> Je höher der Glättungsparameter  $\alpha$  im Intervall zwischen 0 und 1 festgelegt wird, desto schneller sinkt die Bedeutung älterer Werte. Der mit einfacher exponentieller Glättung bestimmte Prognosewert ist definiert als

$$\hat{y}_{T,T+1} = \alpha \cdot y_T + (1 - \alpha) \cdot \hat{y}_{T-1,T}.$$

Die Retourenmenge hängt jedoch unmittelbar von der Menge der versendeten Einheiten ab, da eine Rücksendung an den Versandhändler nur dann erfolgen kann, wenn der Kunde den Artikel oder das Paket vorher erhalten hat. Insofern scheinen **Kausalprognosen** geeigneter, da sie im Gegensatz zu Zeitreihenprognosen derartige funktionale Zusammenhänge berücksichtigen können. Formal kann die bivariate lineare Regression so beschrieben werden:

$$\hat{y}_{T,T+1}(x) = a + b \cdot x_T$$

$x_T$  steht für die unabhängige Variable (nämlich die Anzahl der Bestellungen in Periode  $T$ ), die die abhängige Variable  $y$  (Anzahl der Retouren) erklären soll. Die Parameter  $a$  und  $b$  errechnen sich durch Minimierung der quadrierten Abweichung des Vorhersagewerts vom tatsächlichen Wert in den Vergangenheitsdaten (KQ-Methode).<sup>37</sup> Die Regressionskoeffizienten können in einem festen Turnus (z. B. jährlich) neu berechnet werden, aber auch eine kontinuierliche, quasi rollierende Berechnung anhand der Informationen der letzten  $N$  Perioden ist möglich.

<sup>35</sup> Vgl. für den folgenden Absatz Schröder (2012), S. 22f.

<sup>36</sup> Vgl. für den folgenden Absatz Schröder (2012), S. 25–27.

<sup>37</sup> Vgl. Dreger/Kosfeld/Eckey (2014), S. 24f.



## 2.5 Evaluation von Prognosemodellen

Das Bestimmtheitsmaß ( $R^2$ ) eines linearen Regressionsmodells misst die Anpassungsgüte eines Modells an gegebene (Trainings-)Daten, ist für die Problemstellung der Prognoseevaluation jedoch weniger geeignet, zumal für Zeitreihenprognosen keine vergleichbare Größe existiert. Zielsetzung eines Prognosemodells ist generell eine gute Generalisierungsfähigkeit für neue Daten.<sup>38</sup> Nur der nachträgliche Vergleich der prognostizierten Werte für die abhängige Variable mit den realen Werten eines Evaluations- oder Testdatensatzes kann die Prognosequalität objektiv beurteilen, wenn die Modellspezifikation auf anderen Daten (Trainingsdaten) basiert.<sup>39</sup>

Ein solcher ex-post-Prognosefehler lässt sich auf mehrere Arten berechnen, wobei sich die Kennzahlen hinsichtlich der Berechnung und Beurteilung des Fehlermaßes sowie dessen Interpretationsmöglichkeiten unterscheiden. Zum Vergleich verschiedener Modelle kann es bei geringen Modelldifferenzen sinnvoll sein, mehrere Fehlermaße zu betrachten, um heterogene Perspektiven (z. B. unterschiedliche Ausreißergewichtung) einzubeziehen. Tabelle 1 zeigt einige gängige Beurteilungsmaße im Überblick, die zur Evaluation von numerischen Prognosemodellen geeignet sind, erhebt aber keinen Anspruch auf Vollständigkeit.

Berechnung	Bezeichnung und Kurzerläuterung
$AE = \sum_{i=1}^N  \hat{x}_i - x_i $	Absolute Error: absoluter Prognosefehler
$MAE = \frac{1}{N} \sum_{i=1}^N  \hat{x}_i - x_i $	Mean Absolute Error: mittlerer absoluter Fehler
$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2$	Mean Squared Error: mittlerer quadrierter Fehler
$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2}$	Root Mean Squared Error: Wurzel des mittleren quadrierten Fehlers
$MAPE = \frac{1}{N} \sum_{i=1}^N \left  \frac{\hat{x}_i - x_i}{x_i} \right $	Mean Absolute Percentage Error: mittlerer relativer Fehler

Tabelle 1: Ausgewählte Beurteilungsmaße für Prognosen<sup>40</sup>

Der AE ist grundsätzlich nur dann verwendbar, wenn alle Prognosen auf derselben Beobachtungsanzahl basieren; da seine Einheit nicht normiert ist, ist er nicht relativ interpretierbar.<sup>41</sup> Bei der Prognose einer binären abhängigen Variable (0/1) besteht zwischen MAE und MSE kein Unterschied, während der MAPE wegen der möglichen Division durch Null nicht anwendbar ist.

<sup>38</sup> Vgl. Domingos (2012), S. 80.

<sup>39</sup> Vgl. Hansmann (1983), S. 14.

<sup>40</sup> Quelle: Eigene Darstellung in Anlehnung an Hansmann (1983), S. 15f.; Petersohn (2005), S. 171f.; Küsters (2012), S. 434.

<sup>41</sup> Vgl. für den folgenden Absatz Barrot (2009), S. 548–550; Petersohn (2005), S. 171f.

Vorteile des MAE und des RMSE bei numerischen Größen sind deren einfache Interpretierbarkeit bei der Gegenüberstellung zur prognostizierten Größe, während der MSE aufgrund des mitsamt der Einheit (also z. B. Stück<sup>2</sup>) quadrierten Ergebnisses kaum sinnvoll in Bezug zur Ausgangsgröße gesetzt werden kann. Der MAPE ist aufgrund seiner relativen Perspektive am einfachsten deutbar, besitzt allerdings wie auch der MAE gegenüber dem RMSE eine geringere Ausreißersensitivität.<sup>42</sup>

### 3 Data Mining-Modelle zur Prognose der Retourenentscheidung

Das Kapitel 3 beschäftigt sich mit der Erstellung und Auswertung verschiedener Data Mining-Modelle, die für jeden einzelnen versendeten Artikel bzw. jedes versendete Paket vorhersagen sollen, ob daraus eine Rücksendung entsteht oder ob der Artikel bzw. das Paket beim Kunden verbleibt. Hierzu führt Abschnitt 3.1 in den verwendeten Datensatz ein, auf dessen Basis die Modelle erstellt werden. Abschnitt 3.2 erläutert die Vorgehensweise bei der Modellierung, während die Abschnitte 3.3 und 3.4 die damit erzielbare Prognosegenauigkeit auf Artikel und Paketebene bewerten.

#### 3.1 Details zum Datensatz

Die Datengrundlage der Untersuchung besteht aus historischen Transaktionsdaten ( $n=481092$ ) eines realen Onlineshops der Bekleidungs- und Schuhbranche über ein Jahr. Beim betrachteten Versandhändler handelt es sich um einen Mittelständler mit einem Jahresumsatz von 13,2 Mio. € im Beobachtungszeitraum, bei dem in derselben Zeit Rücksendungen im Wert von 17,8 Mio. € eingingen.<sup>43</sup> Die Rücksendequoten in dieser Zeit liegen für alle Bezugsebenen über 50 % ( $\alpha$ -RQ: 59,85 %;  $\beta$ -RQ: 52,55 %;  $\gamma$ -RQ: 57,38 %). Abbildung 1 zeigt einen deutlichen Zusammenhang zwischen Bestellungen und der Retourenanzahl, während über das ganze Jahr eine relativ konstante, mit eher schwachem Trend behaftete Anzahl an Bestellungen und Retouren erkennbar ist und auch die monatliche  $\beta$ -Retourenquote nur gering, nämlich zwischen 50 % und 55 % schwankt (rechte Skala).

---

<sup>42</sup> Vgl. Küsters (2012), S. 433f.

<sup>43</sup> Vgl. zur Mittelstandsdefinition: Europäische Union (2003), S. 39.

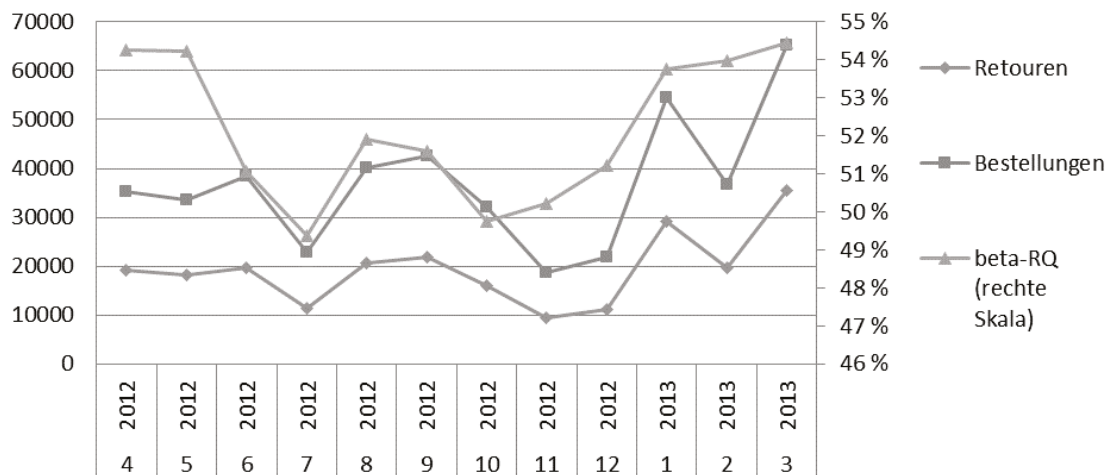


Abbildung 1: Verlauf von Bestellungen, Retouren und beta-Retourenquote in den Transaktionsdaten.<sup>44</sup>

Der Datensatz enthält Informationen über die Transaktion selbst (Verkaufs- und Versanddatum, ID), über das verkaufte Produkt (ID, Größe, Farbe, Hersteller, Preis), zum Kunden (ID, Geschlecht/Anrede, Geburtstag, Bundesland, Erstellungsdatum des Kundenkontos) sowie darüber, ob der Kunde das jeweilige Produkt retourniert hat. Zusätzlich zu den bereits genannten Daten werden die Transaktionen eines direkt anschließenden Monats ( $n=45810$ ) als Testdatensatz verwendet, für die die Rücksendeentscheidung zunächst als unbekannt angenommen wird und die der Evaluation der verschiedenen Modelle dienen. Die vorhergehenden Transaktionsdaten über ein Jahr werden zum Training der Modelle genutzt und daher im Folgenden als Trainingsdaten bezeichnet.

93,5 % der Kunden im betrachteten Datensatz sind weiblich. Tabelle 2 charakterisiert den Kundenkreis und die versendeten Artikel des Händlers näher. Die Ursprungsdaten beinhalten teilweise fehlende, inkonsistent oder unplausible Werte: Eine Bestellung durch eine 112-jährige Kundin ist beispielsweise sehr unwahrscheinlich. Daher ist eine umfassende Vorverarbeitung der Daten nötig. Des Weiteren wurden die vorhandenen Daten durch Datenaggregation (z. B. Erkennung von Auswahlbestellungen) und Integration von Daten aus externen Quellen (bspw. Bevölkerungsdichte der Herkunftsbundesländer der Kunden) erweitert und um zusätzliche Informationen angereichert.

<sup>44</sup> Quelle: Eigene Darstellung.

Variable	N	Mittelwert	SD	Min.	Max.
Kundenalter [Jahre] (ohne fehlende Werte)	53549	47,80	9,94	0	112
Lieferzeit [Tage]	482283	11,25	18,343	0	175
Artikelpreis [Euro]	531170	70,51	44,82	0,00	999,00
Account-Alter [Tage]	531170	476,46	273,661	1	805

Tabelle 2: Lage- und Streuungsmaße verschiedener relevanter Variablen für Test- und Trainingsdatensatz gemeinsam (13 Monate).<sup>45</sup>

### 3.2 Vorgehensweise bei der Modellspezifikation zur Prognose der Retourenentscheidung

Im ersten Schritt ist die Filterung aller Transaktionen mit fehlendem Versand-/Lieferdatum nötig. Diese Artikel können nicht zurückgesendet werden, da sie erst gar nicht beim Kunden ankommen.

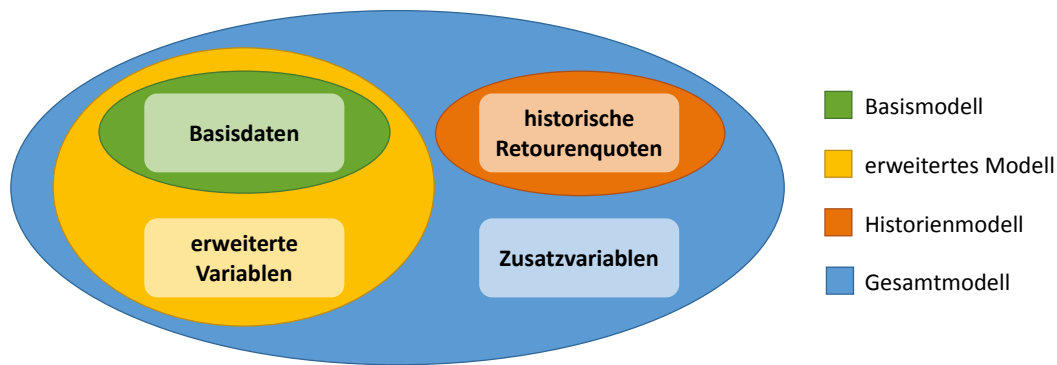
Das Vorgehen für die Modellspezifikation richtet sich auf Artikel- und Paketebene nach folgender Systematik (vgl. dazu Abbildung 2): Zunächst fließen nur die grundlegend im Datensatz enthaltenen metrischen Variablen in die Erstellung des Basismodells ein. Zusammen mit zusätzlichen Erklärungs- bzw. Kontrollvariablen<sup>46</sup> bilden diese ein erweitertes Modell. Das Historienmodell integriert ausschließlich historische Retourenquoten verschiedener Bezugsgrößen (Artikel, Hersteller, Farbe, Kunde, Bundesland), ignoriert aber Informationen über Preise, Auswahlbestellungen etc. Ein Gesamtmodell vereint alle in den Daten verfügbaren potentiellen Einflussgrößen, also auch zusätzliche im Rahmen der Datenvorverarbeitung generierte Faktoren.

Unterschiedliche Algorithmen (binär logistische Regression, neuronale Netze, C5.0-Entscheidungsbäume, Diskriminanzanalyse, Ensemblemodelle<sup>47</sup>) sowie teils verschiedene Vorverarbeitungsschritte erstellen für jede der beschriebenen Datengrundlagen mehrere Modelle, deren Präzision anhand des Testdatensatzes beurteilt wird.

<sup>45</sup> Quelle: Eigene Darstellung.

<sup>46</sup> Erklärungs- und Kontrollvariablen unterscheiden sich letztlich nur nach theoretisch hergeleiteten oder subjektiv plausibilisierten Gesichtspunkten bei der Interpretation des Modells, aus rein technischer Sicht besteht zwischen diesen beiden Gruppen bei der Modellberechnung kein Unterschied.

<sup>47</sup> Die Ensemblemodelle bestehen aus mindestens zwei der genannten Algorithmen oder aus weiteren im SPSS Modeler implementierten Entscheidungsbaumalgorithmen (z. B. CART).

Abbildung 2: Schematische Zuordnung der Variablen zu den Modellen.<sup>48</sup>

### 3.3 Modelle zur Prognose der Retourenentscheidung auf Articlebene

Auf der Articlebene ermöglicht das alle verfügbaren Eingabevariablen umfassende Gesamtmodell die beste Vorhersage: Es prognostiziert zwei Drittel aller Rücksendungen korrekt. Die Elimination erklärungschwacher oder nicht plausibler Variablen kann dieses Gesamtmodell nochmals verbessern und erreicht mit einem Prognosefehler von 14987 falsch klassifizierten Transaktionen (aus 45810) eine relativ hohe Genauigkeit (siehe Tabelle 3).

Datenbasis	Modell	Fehler
Basismodell	C5.0	17869 (39,01 %)
Erweitertes Modell	NN	16682 (36,42 %)
Historienmodell	NN	16080 (35,10 %)
Gesamtmodell	NN	15274 (33,34 %)
Reduziertes Gesamtmodell	DA, NN	14987 (32,71 %)

Tabelle 3: Vergleich der Modellpräzision artikelbasierter Klassifikationsmodelle (n=45810 Artikel).<sup>49</sup>

Der größte Sprung liegt zwischen dem Basismodell und dem erweiterten Modell, offensichtlich spielt also die Identifizierung von Auswahlbestellungen eine wichtige Rolle. Die Güte des Gesamtmodells hebt sich deutlich vom Historienmodell oder vom erweiterten Modell ab. Insofern lohnt sich zugunsten einer hohen Prognosegenauigkeit die Generierung und Aufnahme neuer Variablen. Eine sorgfältige iterative Überprüfung des Variableneinflusses ist zudem unumgänglich, wie das Verbesserungspotential des Gesamtmodells durch Variableneliminierung nachweist. Das reduzierte Gesamtmodell weist zwar eine schlechtere Anpassung an die Trainingsdaten auf, kann unbekannte Daten aber besser vorhersagen, besitzt also eine höhere Generalisierungsfähigkeit.

<sup>48</sup> Quelle: Eigene Darstellung.<sup>49</sup> Quelle: Eigene Darstellung.

Die Performance des Historienmodells verdeutlicht, dass mit geringem Aufwand und ohne Betrachtung bestellspezifischer Informationen über viele Daten gemittelt eine akzeptable Prognose möglich ist.

### 3.4 Modelle zur Prognose der Retourenentscheidung auf Paketebene

Neben der Vorhersage auf Articlebene interessiert für logistische Zwecke besonders die Prognose auf **Paketebene**, welche paketspezifische Einflussgrößen wie Lieferzeit, Paketgröße, und Paketwert als exogene Variablen verwendet. Artikelbezogene Retourenquoten integrieren als zusätzliche Variable Detailinformationen über die im Paket enthaltenen Artikel in die Modelle. Wie auf Articlebene verwenden die Paketmodelle einen sukzessive erweiterten Satz an Prädiktoren, um in der Gesamtevaluation deren Einfluss zu vergleichen und die Modelle unter der Prämisse eines angemessenen Verhältnisses zwischen Komplexität und Güte auszuwerten.

Ein zusätzliches artikelmodellbasiertes Alternativmodell aggregiert die Informationen des besten Artikelmodells für die Paketebene und nutzt dessen Wahrscheinlichkeiten für die jeweilige Entscheidung auf Artikelbasis – Rücksendung ja oder nein – zur Bestimmung der Rücksendewahrscheinlichkeit eines Pakets:

$$p_j(\text{Paket}) = 1 - \prod_{i=1}^N (1 - p_i(\text{Artikel}))$$

Pakete mit einer gemäß dieser Formel berechneten Rücksendewahrscheinlichkeit größer als 50 % klassifiziert das Modell als zu erwartende Retoure.

Im Vergleich (Tabelle 4) schneidet dieses artikelbasierte Alternativmodell am besten ab und ermöglicht zudem eine leicht bessere Prognose als auf Articlebene. Generell ist auf der Paketebene die Vorhersage präziser, was mit der grundsätzlich höheren  $\alpha$ -Retourenquote zusammenhängt: Ein Modell, das *alle* Einheiten als Rücksendung klassifiziert, würde für *Pakete* einen Fehler von etwa 40 % begehen, während es 48 % der *Artikel* falsch zuordnen würde.

Die Ensemble-Modelle, die aus kombinierten Modellen verschiedener Verfahrensklassen bestehen, zeichnen sich durch die besten und stabilsten Prognosen aus. Die Unterschiede zwischen den verschiedenen Datenbasen sind auf Paketebene vergleichsweise gering. Daher empfiehlt es sich aus praktischer Sicht, entweder auf wenig aufwändige Historienmodelle zurückzugreifen oder die auf Articlebene vorgenommene detaillierte Modellierung mittels des Alternativmodells direkt für die Paketvorhersage zu nutzen.

Datenbasis	Modell	Fehler
Basismodell	Ensemble	7215 (32,69 %)
Erweitertes Modell	Ensemble	7004 (31,74 %)
Historienmodell	C5.0	6897 (31,25 %)
Gesamtmodell	Ensemble	6672 (30,23 %)
Verbessertes Gesamtmodell	Ensemble	6544 (29,65 %)
Artikelmodelle	Alternativmodell	6195 (28,07 %)

Tabelle 4: Vergleich der Modellpräzision paketbasierter Klassifikationsmodelle (n=22069 Pakete).<sup>50</sup>

## 4 Evaluation von Retourenmengenprognosen

Abschnitt 4.1 erläutert für die Vorhersage der Artikelgesamtmenge ein zusätzliches naives Prognosemodell, dessen Tauglichkeit für die Mengenvorhersage evaluiert werden soll. Im Anschluss evaluieren die Abschnitt 4.2 und 4.3 die wochenweise Prognosegüte der verschiedenen Verfahren sowohl für Artikel- als auch Paketgesamtmengen.

Um den Nutzen der erstellten Data Mining-Modelle für logistische Planungsprozesse zu verifizieren, vergleicht der Abschnitt 4.4 die Prognose für die Anzahl der zu erwartenden Retouren im Zeitraum eines Monats auf Basis der besten Data Mining-Modelle aus Abschnitt 3.2 mit den unter 4.2 und 4.3 verwendeten Verfahren. Die anschließende Betrachtung auf Einzelproduktebene in Abschnitt 4.5 rundet die Analyse der Retourenmengenprognosen ab.

### 4.1 Naive Retourenmengenprognose

Zusätzlich zu den unter 2.4 erläuterten Verfahrensarten kommt eine dritte Berechnungsmethodik zum Einsatz, die ebenfalls die genannten Kausalbeziehung, also die Abhängigkeit der Retouren von aktuellen Absatzzahlen berücksichtigt und daher bei näherer Betrachtung Ähnlichkeiten mit dem obigen Regressionsmodell aufweist. Die Multiplikation der von den Kunden bestellten Menge innerhalb eines Bezugszeitraums mit der historischen Retourenquote lässt sich als naive Schätzung<sup>51</sup> formal folgendermaßen darstellen:

$$N_{\text{bestellte Einheiten}} \cdot RQ_{\text{historisch}} = N_{\text{retournerierte Einheiten}}$$

Es handelt sich gewissermaßen um ein Regressionsmodell ohne Konstante, bei dem die historische Retourenquote als Steigungsparameter dient. Der zeitliche Bezug

<sup>50</sup> Quelle: Eigene Darstellung.

<sup>51</sup> Vgl. Toktay (2001), S. 4.

dieser Retourenquote ist variabel und kann sich z. B. auf die jeweils letzte Periode, einen exponentiell geglätteten Wert der vergangenen Perioden oder aber einen konstanten Zeitraum, also exemplarisch die Retourenquote des vorherigen Geschäftsjahres erstrecken.

#### 4.2 Artikelbezogene Prognose der Retourenmenge des Versandhändlers

Im Folgenden widmet sich die Untersuchung der Anzahl der retournierten Artikel, wobei eine Woche einer Periode entspricht. Hierfür werden die erläuterten quantitativen Methoden auf die Transaktionsdaten des Schuh- und Bekleidungsversandhändlers angewendet und anhand der Fehlermaße MAPE, MAD und RMSE beurteilt. Dieser Evaluation in Tabelle 5 liegt implizit die Annahme zugrunde, dass Prognosen stets am Ende der Periode erstellt werden und Retouren in der Periode nach Bestellung entstehen.

Verfahren	Parameter	MAPE	MAD	RMSE
GMW	N = 1 Periode	29,87 %	1397	2040
	N = 4 Perioden	36,36 %	1575	2260
	N = 8 Perioden	33,81 %	1403	2079
1EXPO	$\alpha = 0,2$	32,76 %	1366	2062
	$\alpha = 0,5$	31,36 %	1378	2032
	$\alpha = 0,8$	30,44 %	1385	2031
2EXPO	$\alpha = 0,5; \beta = 0,2$	35,07 %	1548	2209
	$\alpha = 0,5; \beta = 0,5$	40,15 %	1781	2499
	$\alpha = 0,5; \beta = 0,8$	45,94 %	2092	2810
Bivariate lineare Regression	fix auf Basis der Perioden 1–8	6,04 %	213	260
	rollierend, N = 4 Perioden	3,69 %	143	187
	rollierend, N = 8 Perioden	3,37 %	123	182
Naives Modell	$\alpha = 0,2$	2,60 %	118	186
	$\alpha = 0,5$	2,36 %	101	152
	$\alpha = 0,8$	2,41 %	102	155
	$\alpha = 1$ (Vorperiodenwert der RQ)	2,50 %	106	162

Tabelle 5: Vergleich der vorhergesagten Artikelretourenmengen. Gütemaße berechnet für 49 Perioden ab Periode 9 (Gesamtzahl der Retouren in diesem Zeitraum: 221968, durchschnittlich 4530 Retouren/Periode).<sup>52</sup>

<sup>52</sup> Quelle: Eigene Darstellung.



Da sich die Prognosegüte der einfachen exponentiellen Glättung mit steigendem  $\alpha$  verbessert und für  $\alpha = 1$  den Bestwert erzielt, ist die Annahme eines konstanten Modells kritisch zu hinterfragen: Ein solches Verhalten deutet nach Meinung der Literatur auf die Existenz eines Trends oder von Saisonalität hin.<sup>53</sup> Aus diesem Grund kommt zusätzlich die zweifache exponentielle Glättung nach Holt zum Einsatz, die einen möglichen Trend berücksichtigt.<sup>54</sup> Den Ergebnissen nach liegt jedoch keine derartige Entwicklung vor, da sich für die zweifache exponentielle Glättung die Prognosen verschlechtern bzw. der Parameter  $\beta = 0$  die besten Prognosen dieses Verfahrens berechnet, was wieder der einfachen exponentiellen Glättung entspricht. Saisonalfaktoren ließen sich nur bei einer mehrjährigen Datenbasis berechnen.

Der gleitende Mittelwert über 1 Periode bzw. die exponentielle Glättung mit Glättungsparameter  $\alpha = 1$  entsprechen dem Wert der Vorperiode als Vorhersagewert, welcher innerhalb der Zeitreihenprognosen noch die beste Prognose darstellt, aber einen durchschnittlichen Fehler von knapp 30 % aufweist. Generell eignen sich die Zeitreihenprognosen somit aufgrund der Nichtberücksichtigung aktueller Absatzzahlen nicht für die Prognose der Retourenmenge.

Die Kausalverfahren schneiden wesentlich besser ab, wobei aufgrund der Einfachheit und wegen der besseren Prognose das naive Verfahren der lineare Regression vorzuziehen ist. Die exponentielle Glättung der Retourenquote mit einem mittleren bis hohen  $\alpha$  kann je nach Schwankungsanfälligkeit der zugrunde liegenden Daten die bestmöglichen Ergebnisse erzielen, in den hier genutzten Daten liefert aber bereits der Verwendung der Retourenquote der Vorperiode ( $\alpha = 1$ ) eine sehr geringe Prognoseabweichung von nur 2,60 %.

#### *4.3 Paketbezogene Prognose der Retourenmenge des Versandhändlers*

Der primäre Nutzen der Mengenvorhersage auf Unternehmensebene liegt in der Vorhersage der Paketmenge, die vom Unternehmen zu bearbeiten ist, da diese Daten bspw. die operative Personaldisposition unterstützen.

---

<sup>53</sup> Vgl. Schröder (2012), S. 30.

<sup>54</sup> Vgl. Schröder (2012), S. 41f.; Holt (2004), S. 5–10.

Verfahren	Parameter	MAPE	MAD	RMSE
GMW	N = 1 Periode	30,70 %	714	1040
	N = 4 Perioden	37,39 %	814	1174
	N = 8 Perioden	36,18 %	759	1093
1EXPO	$\alpha = 0,2$	35,06 %	728	1074
	$\alpha = 0,5$	32,73 %	713	1050
	$\alpha = 0,8$	31,37 %	706	1041
2EXPO	$\alpha = 0,5; \beta = 0,2$	35,80 %	801	1145
	$\alpha = 0,5; \beta = 0,5$	41,59 %	917	1293
	$\alpha = 0,5; \beta = 0,8$	47,89 %	1077	1447
Bivariate lineare Regression	fix auf Basis der Perioden 1–8	5,31 %	113	140
	rollierend, N = 4 Perioden	4,50 %	86	117
	rollierend, N = 8 Perioden	3,48 %	71	95
Naives Modell	$\alpha = 0,2$	2,70 %	56	72
	$\alpha = 0,5$	2,48 %	50	65
	$\alpha = 0,8$	2,39 %	48	65
	$\alpha = 1$ (Vorperiodenwert der RQ)	2,45 %	51	67

Tabelle 6: Vergleich der vorhergesagten Paketretourenmengen. Gütemaße berechnet für 49 Perioden ab Periode 9 (Gesamtzahl der Retouren in diesem Zeitraum: 114921, durchschnittlich 2345 Retouren/Periode).<sup>55</sup>

Die Ergebnisse in Tabelle 6 bestätigen die Erkenntnisse aus der Prognose der Artikelmenge: Für die Paketebene kann wiederum das naive Modell die Mengen am besten abschätzen. Die bivariate lineare Regression kommt ebenfalls zu brauchbaren Ergebnissen, aufgrund des höheren Aufwandes ist aber das naive Modell ohne Glättung oder mit einfacher exponentieller Glättung mit einem mittleren bis hohen Glättungsparameter  $\alpha$  vorzuziehen. Zeitreihenverfahren schneiden etwa eine Zehnerpotenz schlechter ab.

#### 4.4 Vergleich der naiven Gesamtmengenprognosen mit der Mengenprognose des Data Mining-Modells für den Testdatensatz

Da das Data Mining-Modell auf den kompletten Trainingsdaten beruht, ist eine objektive Evaluation mit unbekanntem Daten nur für den Monat des Testdatensatzes

<sup>55</sup> Quelle: Eigene Darstellung.

möglich und erfolgt für die Gesamtretourenanzahl in ebendiesem Monat. Als Vergleichs- bzw. Referenzmodell fungiert das voran beschriebene naive Modell.

Modell	Artikelebene: 25029 retournierte Artikel		Paketebene: 13166 retournierte Pakete	
	Vorhersage	Relative Abweichung	Vorhersage	Relative Abweichung
<b>Data Mining-Modell</b>	28671	15,08 %	14799	12,40 %
<b>Naives Modell</b>	24072	3,82 %	13208	0,32 %

Tabelle 7: Prognosevergleich des Data Mining-Modells und des naiven Modells für die Gesamtretourenmenge eines Monats (Testdaten: April 2013).<sup>56</sup>

Dem naiven Modell in Tabelle 7 liegt die historische Retourenquote gemittelt über das vorige Geschäftsjahr zugrunde. Durch Verwendung der Retourenquote des Vormonats anstelle der des ganzen Jahres beträgt die Abweichung nur 0,34 %, ist aber anfälliger gegenüber kurzzeitigen Schwankungen dieser Quote. Die exponentielle Glättung dieser Retourenquote mit Glättungsparameter  $\alpha = 0,5$  führt zu einem Prognosefehler von 1,55 % und sollte einen guten Kompromiss bezüglich der Reagibilität auf Veränderungen des Rücksendeverhaltens darstellen.

Auf Paketebene ist die Vorhersage etwas präziser als auf Artikelebene: Bei 13166 tatsächlich zurückgesendeten Paketen in den Testdaten liegt der Fehler bei nur 0,32 %, während er bei Verwendung der  $\alpha$ -Retourenquote des Vormonats auf 0,61 % steigt. Bei Verwendung einer monatsweise exponentiell geglätteten Retourenquote mit einem Glättungsparameter von  $\alpha = 0,5$  liegt der Fehler bei nur 0,71 %, für  $\alpha = 0,25$  bei 0,07 %.

Verglichen mit dem besten Data Mining-Modell weicht das naive Modell wesentlich weniger vom wahren Wert ab und eignet sich folglich besser zur Gesamtmengenprognose (vgl. Tabelle 7).

#### 4.5 Prognose der Retourenmenge auf Produktebene

Auf Einzelproduktebene besitzt die Prognose der zu erwartenden Retourenmenge den höchsten praktischen Wert, da diese Zahlen direkt für die Bestellmengen- und Lagerbestandsplanung der einzelnen Produkte im Sortiment nutzbar sind. Vermutlich problematisch ist die Vorhersage bei Artikeln mit niedrigen Verkaufs- und Rücksendezahlen, weshalb der Analysezeitraum den kompletten Monat der Testdaten umfasst, um die für die Prognose nötigen Stückzahlen sicherzustellen. Dennoch stellt dieser Abschnitt auch dar, inwiefern wenig verkaufte und zurückgesendete Artikel prognostizierbar sind.

<sup>56</sup> Quelle: Eigene Darstellung.

Die Mengenprognose des naiven Modells für den Testzeitraum berechnet sich aus der über ein Jahr (entspricht den Trainingsdaten) aggregierten produktbezogenen  $\beta$ -Retourenquote multipliziert mit den Absatzzahlen des auf dieses Jahr folgenden Monats (also in den Testdaten). Die Prognose des Data Mining-Modells entstammt dem besten Artikelmodell, also dem reduzierten Gesamtmodell, für das jeweils auf Produktebene die Summe der vorhergesagten Rücksendungen berechnet wurde.

Abbildung 3 zeigt die höhere Prognosegenauigkeit des naiven Modells sowie den Zusammenhang der mit dem MAPE gemessenen Prognosegüte von der Anzahl retournierter Artikel: Die Retourenmenge selten gekaufter Artikel lässt sich schlechter und folglich nur in längeren Zeiträumen vorhersagen als die von Top-Sellern. Bei dieser ungewichteten Analyse der Artikel schränken mehrere Iterationen die Anzahl der berücksichtigten Produkte anhand deren Bestellhäufigkeit sukzessive ein: Im Analysezeitraum wurden beispielsweise 575 verschiedene Produkte mehr als 5 mal retourniert, während nur 256 Produkte mehr als 25 Rücksendungen verursachten. Artikel ohne Rücksendung bleiben unberücksichtigt, da sich für diese Gruppe keine prozentuale Abweichung vom wahren Wert (Null) bestimmen lässt.

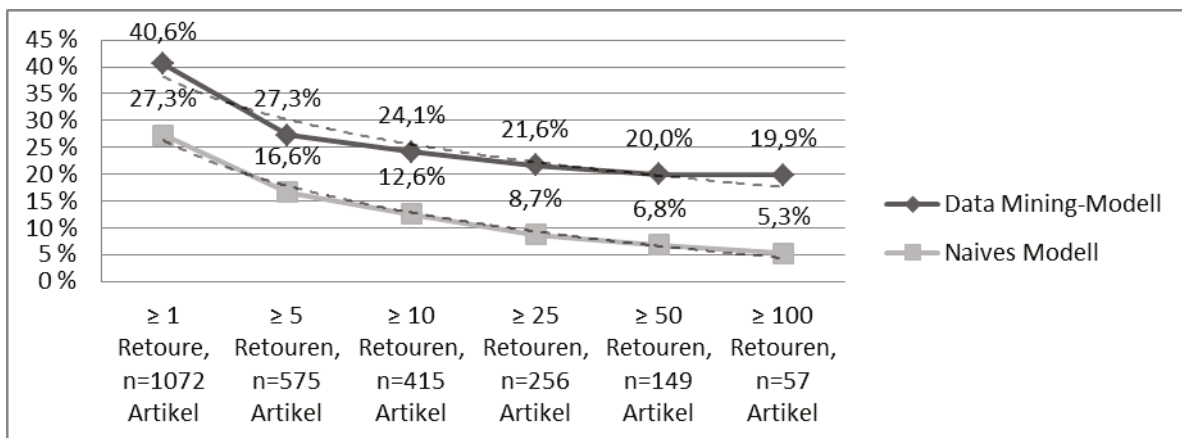


Abbildung 3: Mittlere prozentuale betragsmäßige Abweichung der produktindividuellen Mengenprognose (MAPE) in Abhängigkeit von der Mindestanzahl retournierter Artikel. Datenbasis: 45810 Transaktionen im Zeitraum eines Monats (davon 45378 Transaktionen für Produkte mit mindestens einer Rücksendung).<sup>57</sup>

Durch jede Einschränkung des Umfangs der untersuchten Produkte, bei der nach und nach die selten retournierten Artikel ausgefiltert werden, verbessert sich – nicht ganz überraschend – die Prognose der retournierten Menge. Das naive Modell kann die Retourenmenge bereits bei Produkten mit mehr als 5 Retouren mit einer durchschnittlichen Abweichung von 16,6 % besser vorhersagen als das Data Mining-Modell bei allen Produkten mit mehr als 100 Retouren (19,9 % durchschnittliche Prognoseabweichung).

Eine abschließende Gewichtung aller Produkte mit mindestens einer Rücksendung mit der jeweiligen Verkaufsmenge und die anschließende Berechnung der normierten relativen Abweichung bestätigt dieses Ergebnis nochmals: Eine Abweichung von

<sup>57</sup> Quelle: Eigene Darstellung.

durchschnittlich 13,0 % (SD=0,31) bei der naiven Prognose steht einem mittleren Fehler von 26,6 % (SD=0,34) beim besten Data Mining-Modell gegenüber.

## 5 Fazit

Grundlegende Limitation der Untersuchung ist die Tatsache, dass nur die Daten eines einzelnen Versandhändlers aus der Schuh- und Bekleidungsbranche mit einem eher speziellen Kundenkreis (93,5 % weibliche Kunden eher gehobenen Alters) betrachtet wurden und sich die Erkenntnisse somit nur ansatzweise auf andere Versandhändler oder gar andere Branchen übertragen lassen. Demzufolge sollten die Ergebnisse in Zukunft anhand weiterer Realdatensätze oder mittels Simulation kritisch überprüft werden. Zudem könnte eine solche Untersuchung auch Rücksendezeitpunkte berücksichtigen, die aufgrund der Datenbasis in der vorliegenden Analyse nicht näher betrachtet werden konnten.

Bezugnehmend auf die eingangs aufgeworfenen Fragestellungen lassen sich die Ergebnisse folgendermaßen zusammenfassen:

- Wie präzise können Händler auf Basis von Transaktionsdaten die Rücksendeentscheidung eines Kunden mit Hilfe von Data Mining-Methoden prognostizieren?

Die besten Prognosemodelle klassifizieren 14987 (32,71 %) von 45810 Transaktionen falsch. Bei Berücksichtigung der vorher ausgefilterten 4268 stornierten Artikel beträgt die Gesamtpräzision im Testdatensatz gut 70 %. Ein Paketmodell auf Basis dieser Artikelmodelle kann 15874 (71,93 %) von 22069 Paketen richtig zuordnen. Die Datenaggregation von Artikel- zu Paketebene verbessert also die Prognosegüte leicht.

- Welche Verfahren eignen sich am besten zur Prognose von Retourenmengen?

Die Auswahl des geeigneten Prognoseverfahrens richtet sich nach Kriterien (insbesondere „[...] Genauigkeit, Kosten und Komplexität [...]“<sup>58</sup>), welche meist in komplementärer Beziehung zueinander stehen und folglich eine sorgfältige Priorisierung erfordern.<sup>59</sup> Diese Untersuchung weist nach, dass dieser Zielkonflikt nicht immer gilt, da das naive Modell zur Mengenprognose auf Basis der historischen Retourenquote und der aktuellen Bestellmenge den komplexeren Regressionsmodellen deutlich überlegen ist. Zeitreihenverfahren eignen sich generell eher weniger zur Retourenmengenprognose.

- Ermöglichen die Data Mining-Modelle zur Retourenentscheidung die Ableitung einer präzisen Retourenmengenprognose?

In der Gesamtbetrachtung sind die zur Prognose der Retourenentscheidung verwendeten, eigentlich jedoch zur Vorhersage der Retourenentscheidung erstellten Data

---

<sup>58</sup> Hansmann (1983), S. 141.

<sup>59</sup> Vgl. Hansmann (1983), S. 141f.

Mining-Klassifikationsmodelle weder für die Gesamretourenmenge des Händlers noch für die Abschätzung der produktspezifische Retourenmenge gut geeignet. Für diese Zwecke können jedoch in einer zukünftigen Untersuchung alternative und speziell darauf ausgerichtete numerische Data Mining-Modelle generiert werden, welche möglicherweise eine ähnliche Präzision wie die naive Mengenprognose erzielen können.

Dieses in diesem Beitrag vorgestellte naive Prognoseinstrument empfiehlt sich aufgrund geringer Abweichungen und wegen der intuitiven Handhabung sehr gut für den Einsatz zur Mengenprognose in der Praxis. In einem nächsten Schritt ist zur frühzeitigen Retourenabschätzung eine Verknüpfung mit produktspezifischen Nachfrageprognosen anstelle der verwendeten Bestellmengen denkbar.

## 6 Literaturverzeichnis

- Asdecker, B. (2011): Retourenmanagement – Begriff, Inhalte, Strategien, in: WiSt – Wirtschaftswissenschaftliches Studium, Jg. 40, H. 5, S. 258–262.
- Asdecker, B. (2012): Retourenmanagement, in: Klaus, P./Krieger, W./Krupp, M. (Hrsg.): Gabler Lexikon Logistik. 5. Auflage, Wiesbaden, S. 495.
- Asdecker, B. (2013): Retouren bringen negative Zinsen, in: LOG.Kompass, Jg. 10, H. 4, S. 42–44.
- Asdecker, B. (2014): Retourenmanagement im Versandhandel – Theoretische und empirisch fundierte Gestaltungsalternativen für das Management von Retouren, Bamberg.
- Barrot, C. (2009): Prognosegütemaße, in: Albers, S./Klapper, D./Konradt, U./Walter, A./Wolf, J. (Hrsg.): Methodik der empirischen Forschung. 3. Auflage, Wiesbaden, S. 547–560.
- Cleve, J./Lämmel, U. (2014): Data Mining, München.
- Crone, S. (2010): Neuronale Netze zur Prognose und Disposition im Handel, Wiesbaden.
- Domingos, P. (2012): A few useful things to know about machine learning, in: Communications of the ACM, Jg. 55, H. 10, S. 78–87.
- Dreger, C./Kosfeld, R./Eckey, H. (2014): Ökonometrie. Grundlagen – Methoden – Beispiele. 5. Auflage, Wiesbaden.
- Dziemba, O./Wenzel, E. (2014): #wir: Wie die Digitalisierung unseren Alltag verändert, München.

- Europäische Union (2003): Empfehlung der Kommission vom 6. Mai 2003 betreffend die Definition der Kleinstunternehmen sowie der kleinen und mittleren Unternehmen, [http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=uriserv:OJ.L\\_.2003.124.01.0036.01.DEU](http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=uriserv:OJ.L_.2003.124.01.0036.01.DEU), Stand: 27.06.2015.
- Europäische Union (2011): Richtlinie 2011/83/EU des europäischen Parlaments und Rates vom 25. Oktober 2011, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:304:0064:0088:de:PDF>, Stand: 14.06.2015.
- Feindt, M./Jarke, M. (2014): Interview mit Michael Feindt zum Thema „Präskriptive Big-Data-Analyse“, in: *Wirtschaftsinformatik*, Jg. 56, H. 5, S. 329–331.
- Feindt, M./Sinn, M. (2014): Blue Yonder Big Data Analytics / Big Data Analytics – wie automatisierte Entscheidungen OTTO entlang des gesamten Produktlebenszyklus unterstützen, [http://www.bitkom-bigdata.de/sites/default/files/BDS2014\\_1400%20FeindtProf%20BlueYonder\\_Sinn%20Otto.pdf](http://www.bitkom-bigdata.de/sites/default/files/BDS2014_1400%20FeindtProf%20BlueYonder_Sinn%20Otto.pdf), Stand: 03.03.2015.
- Fidan, N./Ozkok, B. (2013): A Review on Applied Data Mining Techniques to Stock Market Prediction, in: Papajorgji, P./Guimarães, A./Guarracino, M. (Hrsg.): *Enterprise Business Modeling, Optimization Techniques, and Flexible Information Systems*, Hershey, S. 108–126.
- Fromm, S. (2005): *Binäre logistische Regressionsanalyse*, Bamberg.
- Garvin, D. (1993): Building a Learning Organization, in: *Harvard Business Review*, Jg. 71, H. 4, S. 78–91.
- Handelsverband Deutschland (2014a): Online Umsatz nach Sortiment, <http://www.einzelhandel.de/index.php/presse/zahlenfaktengrafiken/item/124658-online-umsatz-nach-sortiment>, Stand: 02.04.2015.
- Handelsverband Deutschland (2014b): E-Commerce-Umsätze, <http://www.einzelhandel.de/index.php/presse/zahlenfaktengrafiken/item/110185-e-commerce-umsaetze>, Stand: 03.04.2015.
- Hansmann, K. (1983): *Kurzlehrbuch Prognoseverfahren: mit Aufgaben und Lösungen*, Wiesbaden.
- Holt, C. (2004): Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages, Pittsburg, in: *International Journal of Forecasting*, Jg. 20, H. 1, S. 5–10.
- Jahn, M. (2013): Online versus stationär? Lieber komplementär. Download nach Anmeldung unter <http://www.gfk.com/de/loesungen/geomarketing/Seiten/Landingpage/White-Paper-Handel.aspx>, Stand: 05.02.2015.
- Kaplan, R./Norton, D. (1996): *The balanced scorecard: translating strategy into action*, Boston/Massachusetts.

- Kuß, A./Wildner, R./Kreis, H. (2014): Marktforschung. Grundlagen der Datenerhebung und Datenanalyse. 5. Auflage, Wiesbaden.
- Küsters, U. (2001): Data Mining Methoden: Einordnung und Überblick, in: Hippner, H./Küsters, U./Meyer, M./Wilde, K. (Hrsg.): Handbuch Data Mining im Marketing, Braunschweig/Wiesbaden, S. 95–130.
- Küsters, U. (2012): Evaluation, Kombination und Auswahl betriebswirtschaftlicher Prognoseverfahren, in: Mertens, P./Rässler, S. (Hrsg.): Prognoserechnung. 7. Auflage, Heidelberg/Dordrecht/London/New York, S. 423–467.
- Loebich, M./BearingPoint (2014): Überbestände und Retourenwahnsinn machen Onlinehandel unprofitabel, <http://www.bearingpoint.com/de-de/7-9697/ueberbestaende-und-retourenwahnsinn-machen-onlinehandel-unprofitabel>, Stand: 01.10.2015.
- Petersohn, H. (2005): Data Mining. Verfahren, Prozesse, Anwendungsarchitektur, München.
- Ratner, B. (2012): Statistical and Machine-Learning Data Mining. 2. Auflage, Boca Raton.
- Rogers, D./Lambert, D./Croxtton, K./Garcia-Dastugue, S. (2002): The Returns Management Process, in: The International Journal of Logistics Management, Jg. 13, H. 2, S. 1–18.
- Rosentreter, J. (1977): Prognosen in der industriellen Planung, Frankfurt/Zürich.
- Schröder, M. (2012): Einführung in die kurzfristige Zeitreihenprognose und Vergleich der einzelnen Verfahren, in: Mertens, P./Rässler, S. (Hrsg.): Prognoserechnung. 7. Auflage, Heidelberg/Dordrecht/London/New York, S. 11–45.
- Toktay, B. (2001): Forecasting Product Returns. INSEAD working paper, <http://www.prism.gatech.edu/~bt71/articles/forecasting.pdf>, Stand: 16.05.2015.
- Weber, J./Wallenburg, C. (2010): Logistik- und Supply Chain Controlling. 6. Auflage, Stuttgart.
- Werner, H. (2013): Supply Chain Management. Grundlagen, Strategien, Instrumente und Controlling. 5. Auflage, Wiesbaden.