

Empirische Analyse der Produktbewertungen auf Amazon

Bachelorarbeit

eingereicht bei
Jun.-Prof. Dr. Jochen Reiner
Juniorprofessur Marketing

Fachbereich Wirtschaftswissenschaften
Johann Wolfgang Goethe-Universität
Frankfurt am Main

von
Malik Stromberg

Pfarrer-Papon-Str. 3
64546 Mörfelden-Walldorf
0049 178 1490197
E-Mail: malik.stromberg@arcor.de

Studienrichtung: Wirtschaftswissenschaften
9. Fachsemester
Matrikelnummer: 5884821

Inhaltsverzeichnis

Abbildungsverzeichnis	IV
Tabellenverzeichnis	V
Abkürzungsverzeichnis	VI
Symbolverzeichnis	VII
1 Problemstellung	1
2 Vertrauen in elektronischen Märkten	3
3 Kundenrezensionssysteme	5
3.1 Anforderungen an Rezensionssysteme	6
3.1.1 Kundenseite	7
3.1.2 Verkäuferseite	9
3.2 Rezensionssysteme von Amazon	10
3.3 Bedeutung einer zusammenfassenden Gesamtbewertung	11
4 Algorithmen zur Bewertung und Sortierung von Elementen	12
4.1 Das arithmetische Mittel	12
4.2 Der Histos-Algorithmus (nach Zacharia et al., 2000)	13
4.3 Der Sporas-Algorithmus (nach Zacharia et al., 2000)	14
4.4 Korrelationsbasierter Bewertungsalgorithmus (nach Zhou et al., 2011)	15
5 Konzeptioneller Rahmen	17
6 Design der empirischen Studie	19
6.1 Stützvektormaschinen (SVM)	21
6.2 Elastische Netze	22
6.3 Zufallswälder	24
6.4 Gradient Boosting Maschine (GBM)	25
7 Datensatz	26
7.1 Ursprünglicher Datensatz	27
7.1.1 Strukturelle Veränderung des Datensatzes	27
7.1.2 Ergänzungen im Datensatz	27
7.2 Datensatz der minimalen und maximalen Bewertungen	28
7.2.1 Strukturelle Veränderungen des Datensatzes	28
7.2.2 Ergänzungen im Datensatz	28
7.3 Finaler Datensatz	28

8 Durchführung und Ergebnisse.....	30
8.1 Prüfung des Algorithmus auf das arithmetische Mittel.....	30
8.2 Trainieren der verschiedenen Modelle des maschinellen Lernens.....	31
8.3 Das Zufallswald-Modell.....	33
9 Diskussion	35
10 Limitationen.....	37
11 Weiterführende Forschungsansätze	38
12 Fazit	39
Anhang	41
Literaturverzeichnis.....	57
Internetquellenverzeichnis	61
Ehrenwörtliche Erklärung	62

Abbildungsverzeichnis

Abbildung 1: Entscheidungsbaum im Online-Handel (nach Bolton et al., 2004)	4
Abbildung 2: Einflussfaktoren der wahrgenommenen Rezensionsnützlichkeit	9
Abbildung 3: Aufbereitung häufig verwendeter Stichworte auf Amazon	10
Abbildung 4: Zusammenfassende Rezensionsstatistik auf Amazon.....	10
Abbildung 5: Reputationsnetzwerk (nach Zacharia et al., 2000).....	14
Abbildung 6: Konzeptioneller Rahmen der Arbeit	18
Abbildung 7: Abweichung der Gesamtbewertung vom arithmetischen Mittel der Einzelbewertungen auf Amazon.....	30
Abbildung 8: RMSE mit 95%-Konfidenzintervall der Modelle mit den besten Einstellungsparametern	32
Abbildung 9: Wichtigkeiten der unabhängigen Variablen im Zufallswald-Modell	34

Tabellenverzeichnis

Tabelle 1: Deskriptive Statistiken ausgewählter Variablen des finalen Datensatzes.....	29
Tabelle 2: Optimale Einstellungsparameter der trainierten Modelle	32
Tabelle 3: Modellgesamtergebnis des Zufallswaldes	33
Tabelle 4: Datenstruktur der verwendeten Variablen	41
Tabelle 5: Einstellungen und Ergebnisse der GBM-Probelaufe	42
Tabelle 6: Einstellungen und Ergebnisse der RF-Probelaufe	42
Tabelle 7: Einstellungen und Ergebnisse der RF-Probelaufe mit feinem Gitter.....	42
Tabelle 8: Einstellungen und Ergebnisse der EN-Probelaufe	43
Tabelle 9: Einstellungen und Ergebnisse der SVM-Probelaufe.....	44
Tabelle 10: Einstellungen und Ergebnisse der GBM-Probelaufe mit feinem Gitter	45
Tabelle 11: RMSE der jeweils besten Modelle der verschiedenen Ansätze.....	56
Tabelle 12: Wichtigkeiten der UV im finalen RF-Modell	56

Abkürzungsverzeichnis

ASIN	Amazon-Standard-Identifikationsnummer
AV	abhängige Variable
EN	Elastische Netze
GBM	Gradient Boosting Maschine
IncNodePurity	Anstieg der Knotenreinheit
MAE	mittlerer absoluter Fehler (engl.: mean absolute error)
MSE	mittlerer quadratischer Fehler (engl.: mean squared error)
n.minobsinnode	Mindestanzahl der Beobachtungen in abschließenden Knoten
n.trees	Anzahl der Entscheidungsbäume
OLS	kleinste Quadrate (engl.: ordinary least squares)
RAM	Direktzugriffsspeicher (engl.: random access memory)
RF	Zufallswald (engl.: Random Forest)
RMSE	Wurzel der mittleren Fehlerquadratsumme (engl.: root mean square error)
SSR	Summe der quadrierten Residuen (engl.: sum of squared residuals)
SVM	Stützvektor-Maschine
UV	unabhängige Variable
%IncMSE	prozentualer Anstieg des mittleren quadratischen Fehlers

Symbolverzeichnis

R_i :	Gesamtbewertung des Bewertungselementes i
n_i :	Anzahl vorhandener Bewertungen für Element i
r_{ij} :	Bewertung des Teilnehmers j für Element i
$R_k(d)$:	individuelle Reputation des k -ten Teilnehmers mit Distanz d aus Sicht von T_0
d :	Distanz zum Ausgangsnutzer
D :	höchste erreichbare Reputation
$W_{jk}(d)$:	Bewertung des j -ten Teilnehmers für k -ten Teilnehmer mit Distanz d
R_t :	alter Reputationswert
R_{t+1} :	neuer Reputationswert
R_{t+1}^{other} :	Reputationswert des Bewertenden
t :	Anzahl der bereits erhaltenen Bewertungen
θ :	Zeitgewichtung
W_{t+1} :	Bewertung der neuen Transaktion
D :	höchste erreichbare Reputation
τ :	Beschleunigungsfaktor
Corr_j :	Korrelation der Bewertungen des j -ten Teilnehmers mit den jeweiligen Produkt-gesamtbewertungen
k_{uj} :	Anzahl der abgegebenen Bewertungen des j -ten Teilnehmers
$r_{j\alpha}$:	abgegebene Bewertung des j -ten Teilnehmers für Objekt α
\bar{r}_j :	Erwartungswert der Bewertungen des j -ten Teilnehmers
\bar{Q}_{0j} :	Erwartungswert der geschätzten Gesamtbewertungen der vom j -ten Teilnehmer bewerteten Produkte
$Q_{0\alpha}$:	aggregierte geschätzte Gesamtbewertung des Objektes α
σ_v :	Standardabweichung des Vektors v
O_j :	Menge der Objekte, die vom j -ten Teilnehmer bewertet wurden
U_α :	Menge der Nutzer, die Objekt α bewertet haben
C_{uj} :	Reputation (Gewichtung) des j -ten Teilnehmers
β_l :	l -ter Steigungsparameter
SSR :	Summe der quadrierten Residuen

λ_1 :	exogener Faktor
λ_2 :	exogener Faktor
m :	Anzahl der Steigungsfaktoren
n :	Stichprobengröße
nx_add	relative Anzahl der im Zeitintervall hinzugekommenen Rezensionen mit Bewertungshöhe x
nx_del	relative Anzahl der im Zeitintervall entfernten Rezensionen mit Bewertungshöhe x
tx_add	relatives Durchschnittsalter der im Zeitintervall hinzugekommenen Rezensionen mit Bewertungshöhe x
nx_del	relatives Durchschnittsalter der Löschung von im Zeitintervall entfernten Rezensionen mit Bewertungshöhe x

1 Problemstellung

Der Wegfall direkter sozialer Interaktion im Einzelhandel hat Raum für eine Transformation hinsichtlich der Produktqualitätsvermittlung gegeben. Kundenrezensionen können unter der Voraussetzung, dass sie von Betreibern von Online-Plattformen ausgewertet und aufbereitet werden, für alle Marktteilnehmer von Vorteil sein. Amazon hat es sich als größter Online-Anbieter weltweit (Statista.com [1]) zur Aufgabe gemacht, einen neuen Auswertungsalgorithmus zu entwickeln und hat ihn Anfang des Jahres 2018 im eigenen Rezensionssystem aktiviert. Diese Bachelorarbeit zeigt neue Elemente des Algorithmus auf und analysiert, welcher wesentliche Einflussfaktor nun auf die Gesamtbewertung der Produkte bei Amazon wirkt.

Um die Anzahl der Transaktionen und damit den Umsatz zu maximieren, ist es als Online-Anbieter zum einen nötig, für einen reibungslosen Ablauf und zum anderen für möglichst geringe Hemmschwellen für Käufer und Verkäufer auf der Plattform zu sorgen. Dafür gilt es, das Vertrauen der Käufer in die Plattform und Anbieter, mit denen sie nicht persönlich in Kontakt treten, zu gewinnen (Smith/Brynjolffson, 2001). Ein Instrument, um dies zu gewährleisten, ist ein Kundenrezensionssystem. Kunden können an dieser Stelle das Produkt, den Verkäufer, sowie die Kaufabwicklung bewerten und kommentieren. Damit können sie eine detaillierte Kauf- oder Nichtkaufempfehlung an andere potenzielle Kunden abgeben und zukünftiges Käuferverhalten beeinflussen. Dieses System kann Transparenz schaffen und hilft dadurch den Kunden, sich für das Produkt und den Anbieter zu entscheiden, die ihnen den größten Nutzen stiften können. Verkäufer werden mit dem System unter Druck gesetzt, die Qualität der Produkte und des Versandes zu sichern (Akerlof, 1970).

Hierbei ist die einseitige Machtverteilung zugunsten der Kunden auffällig. Mit entsprechend vielen negativen Bewertungen kann dafür gesorgt werden, dass ein Verkäufer bei potenziellen Kunden als nicht vertrauenswürdig eingestuft wird. Es besteht die Möglichkeit auf Kundenseite, den Markt zu manipulieren. Die Tragweite dieser Möglichkeit wird dann deutlich, wenn die Marktteilnehmer nicht einer Marktseite eindeutig zuordbar sind. Oft können Verkäufer auch als Kunden auf einer Plattform agieren. Dadurch entwickelt sich ein Anreiz für Verkäufer, die Produkte der Konkurrenten negativ oder die eigenen Produkte positiv manipulativ zu bewerten, um den eigenen Absatz zu steigern.

Kundenrezensionen enthalten wichtige Informationen für beide Marktseiten. Für die Kunden wird Transparenz bezüglich der Vertrauenswürdigkeit des Anbieters geschaffen und die Verkäufer erhalten Feedback zu ihren Leistungen. Um durch Suchkosten der Kunden verursachtes Marktversagen zu vermeiden, ist es nötig, die Menge an Rezensionen sinnvoll zusammenzufassen und aufzubereiten (Bakos, 1997). Oft werden für jedes Produkt das arithmetische Mittel und die Verteilung der Rezensionen angegeben. Das arithmetische Mittel ist anfällig für die Auswirkungen manipulierter Bewertungen, da jede manipulierte Rezension mit dem gleichen Gewicht wie jede andere Rezension in die Zusammenfassung eingeht. In Verbindung mit dem Manipulationsanreiz ergibt sich ein Risiko durch verzerrte Bewertungen, dem sich konkurrierende Verkäufer auf der Plattform aussetzen (Chevalier/Mayzlin, 2006). Zusätzlich haben Rezensionen in diesem Bewertungssystem kein Verfallsdatum. Als Konsequenz hängen Anbietern nach Produkt- und Leistungsverbesserungen die negativen Bewertungen der Vergangenheit nach. Auch auf Käuferseite ergibt sich ein Risiko, wenn sich manipulierte Rezensionen nicht von anderen Rezensionen unterscheiden lassen (Jin/Kato, 2006). Die Käufer laufen Gefahr, sich aufgrund der verzerrten Bewertung nicht für das Produkt zu entscheiden, welches ihren individuellen Nutzen maximiert.

Es liegt im Interesse der Online-Plattformen, diese Risiken zugunsten der Anzahl der Transaktionen zu minimieren (Resnick/Zeckhauser, 2015). Ein Weg, das zu erreichen, ist eine Veränderung der Zusammenfassung und Aufbereitung der Rezensionen. Die genannten Risiken ließen sich mithilfe eines manipulationsresistenten Algorithmus einschränken. Dazu ist es für die Online-Plattform notwendig, zusätzliche Daten der Rezensenten, die über die jeweilige Bewertung hinausgehen, zu sammeln und diese angemessen in die Gesamtbewertungen der Produkte einfließen zu lassen.

Dieser Bachelorarbeit liegen Daten zugrunde, die seit Januar 2018 eine Abweichung der Gesamtbewertungen der Produkte auf Amazon vom arithmetischen Mittel der zugehörigen einzelnen Bewertungen aufzeigen. Durch diesen Sachverhalt liegt die Vermutung nahe, dass Amazon einen neuen Algorithmus entwickelt hat, der die Gesamtbewertung auf neue Weise berechnet. In dieser Arbeit wird der Algorithmus in seine Bestandteile aufgeschlüsselt. Es wird geprüft, ob die Gewichtung einer Rezension in der Gesamtbewertung abhängig von ihrem Alter ist. Es wird am Beispiel von Amazon gezeigt, wie das Transaktionsrisiko auf

Online-Plattformen gesenkt und die Effizienz von Bewertungssystemen gesteigert werden kann.

Andere Online-Plattformen werden mit den gewonnenen Erkenntnissen angeregt, einen vom arithmetischen Mittel abweichenden Algorithmus zur Berechnung der Gesamtbewertung der Produkte zu entwickeln. Hierzu werden Elemente aufgezeigt, die in diesem Algorithmus eine Rolle spielen können. Das Ziel einer derartigen Veränderung ist es, das Transaktionsrisiko auf Online-Plattformen zu senken und damit Transaktionen zu fördern (Ba/Pavlou, 2002).

Die in dieser Bachelorarbeit vorgenommene technische Analyse der Zusammensetzung von Produktgesamtbewertungen kann ein Leitfaden für die Entscheidungsfindung potenzieller Kunden sein, wie viel Vertrauen in die Bewertung zu setzen ist. Diese Entscheidung hat zur Folge, dass Kunden dazu in der Lage sind, bewusstere Kaufentscheidungen zu treffen und damit ihren gewonnenen Produktnutzen zu maximieren.

Verkäufern auf Amazon wird mit den Erkenntnissen dieser Arbeit Sicherheit für faire Bewertungen ihrer Produkte gegeben. Sie können sich mit der Funktionsweise des Algorithmus vertraut machen und dieses Wissen in der Strategieentwicklung verwenden. Gleichzeitig zeigt diese Arbeit Ansatzpunkte auf, wie manipulative Bewertungen verändert werden können, um sich mit starkem Gewicht in der Gesamtbewertung auszuwirken.

Zunächst gibt diese Bachelorarbeit in den ersten Kapiteln einen Überblick über die besonderen Herausforderungen des Online-Handels. Dabei wird vor allem auf das Vertrauensproblem bei Internettransaktionen eingegangen. Anschließend wird erörtert, inwiefern Rezensionssysteme zur Lösung dieses Problems beitragen können. Bevor Hypothesen über die Elemente des Amazon-Algorithmus aufgestellt werden, werden verschiedene Arten von Bewertungssystemen dargestellt. Daraufhin werden die Methoden erklärt, mit denen die zuvor aufgestellten Hypothesen überprüft werden. Nachdem die Analysen durchgeführt wurden, werden die Ergebnisse dargestellt und diskutiert.

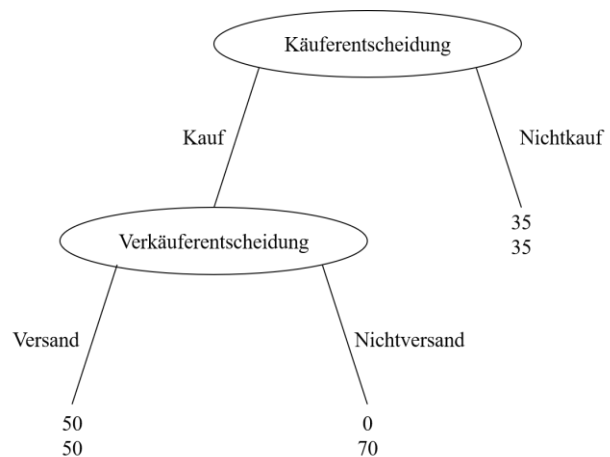
2 Vertrauen in elektronischen Märkten

Neben Kauf- und Verkaufswille stellt Vertrauen eine wichtige Säule eines jeden vollzogenen Handels dar. Das Vertrauen spiegelt sich in informellen und ungeschriebenen Garantien

innerhalb von Transaktionen, die sich über die Zeit entwickelt haben, wider (Akerlof, 1970). Im stationären Handel kann Vertrauen schnell aufgebaut werden. Das Produkt ist gegenwärtig und der Verkäufer tritt als Person auf, mit der der Käufer interagieren kann. Das Ziel eines Verkaufsgesprächs ist es, eine vertraute Verbindung zwischen Verkäufer und Kunden aufzubauen. Der Verkäufer lernt die Bedürfnisse seines Kunden kennen und kann ihn, falls nötig, von seinem Fachwissen überzeugen. Anschließend erhält der Kunde individuelle Produktvorschläge. Dieser Prozess baut Vertrauen auf, auf dessen Grundlage der Kaufabschluss erfolgt (Swan/Nolan, 1985). Der Verkäufer stellt als Vermittler im Vertrauensaufbau ein entscheidendes Element des Kaufprozesses dar. Ein Online-Anbieter kann auf diese direkte menschliche Interaktion nicht zurückgreifen, was ein Vertrauensdefizit in elektronischen Märkten bedingt. Zusätzlich hat der Kunde im stationären Handel die Möglichkeit, sich physisch von der Qualität der angebotenen Produkte zu überzeugen. Das Produkt kann vor dem Kauf hinsichtlich seiner Funktionsweise und -fähigkeit geprüft werden. Auch dieses Element der Vertrauensbildung entfällt bei Online-Anbietern. Es resultieren Informationsasymmetrien, denn hinter Beschreibungen und Bildern birgt sich eine große wahrgenommene Qualitätsvarianz für potenzielle Kunden. Daher bildet sich Unsicherheit auf Käuferseite, die sich in ihrer Zahlungsbereitschaft niederschlägt (Ba/Pavlou, 2002).

Im Internet tritt ein Betrugsfall zwölfmal häufiger auf als im stationären Handel (GartnerG2, 2002 nach Bolton et al., 2004). Mithilfe eines Entscheidungsbaums, wie in Abbildung 1, lässt sich dieses weitere grundlegende Vertrauensproblem in elektronischen Märkten veranschaulichen. Hat der Käufer seine Kaufentscheidung getroffen, lohnt es sich für den Verkäufer finanziell nicht, das Produkt zu versenden (Bolton et al., 2004). Unter der Annahme der vollständigen Transparenz der Auszahlungen entscheidet sich der Käufer in dieser spieltheoretischen Argumentation immer dagegen, das Produkt zu kaufen. Online-Plattformen sind demnach nicht dazu in

Abbildung 1: Entscheidungsbaum im Online-Handel (nach Bolton et al., 2004)



der Lage, das gleiche Maß an Vertrauen zu liefern wie der stationäre Handel (Smith/Brynjolfsson, 2001).

Interessierte Käufer suchen nach Informationen, um ihr Risiko beim Kaufen zu minimieren. Sind diese im großen Maße vorhanden, kann ausreichend Vertrauen aufgebaut werden, sodass es zum Kaufabschluss kommt (Ba/Pavlou, 2002). Dabei werden Informationen zum Anbieter, zum Produkt und, falls beteiligt, zur Online-Plattform herangezogen. Sollten die Informationen über die Qualität des Produktes jedoch mangelhaft sein, wird diese mithilfe der Informationen über den Anbieter approximiert (Resnick et al., 2000). Markenattribute spielen hierbei eine entscheidende Rolle, denn das Markenimage des Anbieters, der Plattform und des Produktes wird auf die jeweilige Glaubwürdigkeit projiziert (Smith/Brynjolfsson, 2001). Um möglichst viele Informationen zusammenzutragen, werden unterschiedliche Quellen verwendet, die mit verschiedenen individuell gesetzten Gewichtungen in den Entscheidungsprozess eingehen (Chen et al., 2008). Dazu gehören soziale Netzwerke, Online-Gemeinschaften und Bewertungsinternetseiten. Unternehmenseigene Online-Gemeinschaften können daher eine wichtige strategische Komponente im Vertrauensaufbau zum Kunden sein. Dabei sollte das Unternehmen Vertrauensaufbau, Kostenminimierung und Kontrolle des Einflusses der Gemeinschaft als Ziele bei der Einführung eines solchen Instruments verfolgen (Walden, 2000).

3 Kundenrezensionssysteme

Es liegt im Interesse von Betreibern von Online-Plattformen, ihr Vertrauensdefizit zugunsten der Markteffizienz zu schließen. Sind die Suchkosten der Käufer zu hoch, kommt es zum Versagen des elektronischen Marktes (Bakos, 1997). Dementsprechend gilt es, die Käufer im Informationssuchprozess zu unterstützen. Potenzielle Käufer besuchen Bewertungsinternetseiten, um bessere Entscheidungen mit weniger Aufwand zu treffen (Dabholkar, 2006). Allerdings kann der Suchprozess unterstützt werden, indem die gesuchten Informationen auf der eigenen Internetseite aufbereitet zur Verfügung gestellt werden. Dieses Informationsangebot erfüllt seinen Zweck, wenn die bereitgestellten Informationen glaubwürdig sind. Um dies zu gewährleisten, greifen viele Plattformen auf ein Rezensionssystem zurück. Innerhalb dieses Systems können Besucher der Internetseite Bewertungen und Beschreibungen des Produktes abgeben und somit unabhängige Informationen für zukünftige Käufer

bereitstellen. Das Rezensionssystem kann dazu in vielerlei Hinsicht nach den Interessen des Betreibers an die Plattform angepasst werden (Dellarocas, 2003). Zunächst kann das Design des Mechanismus konfiguriert und die Anzahl der zu bewertenden Parameter festgelegt werden. Der Plattformbetreiber kann zudem bestimmen, welche Informationen auf welche Art und Weise den Nutzern zur Verfügung gestellt werden. Zusätzlich können Anreize für das Abgeben einer Rezension gesetzt werden. Dabei funktioniert das System, wenn beide Marktseiten glauben, dass es funktioniert (Resnick/Zeckhauser, 2015). Wenn potenzielle Kunden, die auf Informationssuche sind, glauben, den bereitgestellten Informationen vertrauen zu können, wirkt dieses Vertrauen transaktionsfördernd (Hawes et al., 1989). Auf der anderen Marktseite laufen Anbieter Gefahr, für nicht vertrauenswürdiges Verhalten bestraft zu werden. Dadurch wird ein Anreiz für vertrauenswürdiges Verhalten gesetzt. Djawadi et al. (2018) zeigten, dass Anbieter durch Rezensionssysteme die dominante Strategie haben, eine gute Reputation nicht auszunutzen. Das heißt, bietet ein Verkäufer ein Produkt mit hoher Qualität an und erhält dadurch eine gute Bewertung, ist es nicht profitmaximierend, von der angebotenen Qualität zugunsten der Kosten abzuweichen, da zukünftige Gewinne aufgrund der verminderten Informationsasymmetrien geschmälert würden. Darüber hinaus konnten Djawadi et al. (2018) zeigen, dass Anbieter in der Praxis tatsächlich diese dominante Strategie verfolgen. Zusätzlich profitiert eine Plattform von einem Rezensionssystem. Die Verfügbarkeit von Rezensionen erhöht die Verweildauer von Kunden auf der Internetseite und Besucher nehmen die Seite als hilfreich wahr (Kumar/Benbasat, 2006). Dadurch liegt es im Interesse von Plattformbetreibern, ein Anreizsystem für das Verfassen von Rezensionen zu implementieren.

Die Aufgabe des Plattformbetreibers, das Rezensionssystem in seiner Funktionsfähigkeit, bei der die Qualität ein essenzieller Bestandteil ist, aufrecht zu erhalten, gilt es nicht zu unterschätzen. Aufgrund der geringen Identitätswechselkosten im Internet wird eine strategische Manipulation des Systems möglich (Jin/Kato, 2006). Damit das System zweckmäßig bleibt, müssen daher nicht vertrauenswürdige Teilnehmer identifiziert werden (Resnick/Zeckhauser, 2015).

3.1 Anforderungen an Rezensionssysteme

Ziel von Kundenrezensionssystemen ist es, Vertrauen in elektronischen Märkten zu schaffen. Sie setzen am Informationssuchprozess potenzieller Kunden an, indem Erfahrungen

anderer Kunden veröffentlicht werden. Dabei hat ein solches System relevante Aspekte beider Marktseiten zu erfüllen. Kunden gilt es, effizient Produkt- und Anbieterinformationen zur Verfügung zu stellen, um die mögliche Qualitätsvarianz und die wahrgenommene Wahrscheinlichkeit eines Betruges auf Verkäuferseite gering zu halten. Auf der anderen Seite geht es darum, den Anbietern ein faires Feedback garantieren zu können, um sie zu kooperativem Verhalten anzuregen und gleichzeitig vor Manipulation zu schützen. In diesem Abschnitt steht das Hervorheben von Eigenschaften von Rezensionen, die mit diesen Anforderungen zusammenhängen, im Fokus der Analyse. Mithilfe dieser Eigenschaften können später Hypothesen über die enthaltenen Elemente im Gesamtbewertungsalgorithmus anhand des ausgewählten Beispiels von Amazon aufgestellt werden.

3.1.1 Kundenseite

In Summe ergeben alle Kundenrezensionen die Masse an Gesamtinformationen, die anderen potenziellen Kunden durch ein Rezensionssystem zur Verfügung gestellt werden kann. Durch die inzwischen gewonnene Popularität solcher Systeme wird eine beträchtliche Masse an Informationen generiert. Es ergibt sich ein Austauschverhältnis zwischen Suchkosten und Informationsgewinn, dem sich die potenziellen Kunden bewusst sind (Stigler, 1961). Um die Effizienz von Rezensionssystemen sicherzustellen, ändert sich der strategische Fokus der Plattformbetreiber. Dieser liegt nicht mehr nur auf der bloßen Bereitstellung von nutzer-generierten Rezensionen, sondern auf der kundengerechten Aufbereitung der gewonnenen Informationen. Rezensionen stellen eine quasi-soziale Interaktion zwischen Nutzern dar, in dessen Folge sich die Nutzerpräferenzen entwickeln und verändern. Die Präferenzentwicklung ist ausschlaggebend für den erhaltenen Mehrwert für den Kunden und die damit verbundene wahrgenommene Nützlichkeit der Rezension (Gretzel/Fesenmaier, 2006). Jene Informationen, die den Nutzern nützlich erscheinen, sind im Sinne der Einsparung von Suchkosten hervorzuheben (Mudambi/Schuff, 2010). Dabei sind Verfasserqualität und Inhaltsqualität der Rezension die essenziellen Einflussfaktoren für Nützlichkeit (Chen et al., 2008).

Eine Kundenrezension umfasst diejenigen Informationen, die von einem einzelnen Plattformnutzer für ein bestimmtes Produkt zur Verfügung gestellt werden. Bei Amazon gehören dazu in jedem Fall eine Bewertung auf einer Skala von eins (negativ) bis fünf (positiv), sowie optional Text, Bild und Video. Da die Untersuchung des Rezensionssystems von Amazon

im Mittelpunkt dieser Arbeit steht, wird im weiteren Verlauf des Kapitels auf Rezensionen dieser beschriebenen Form eingegangen.

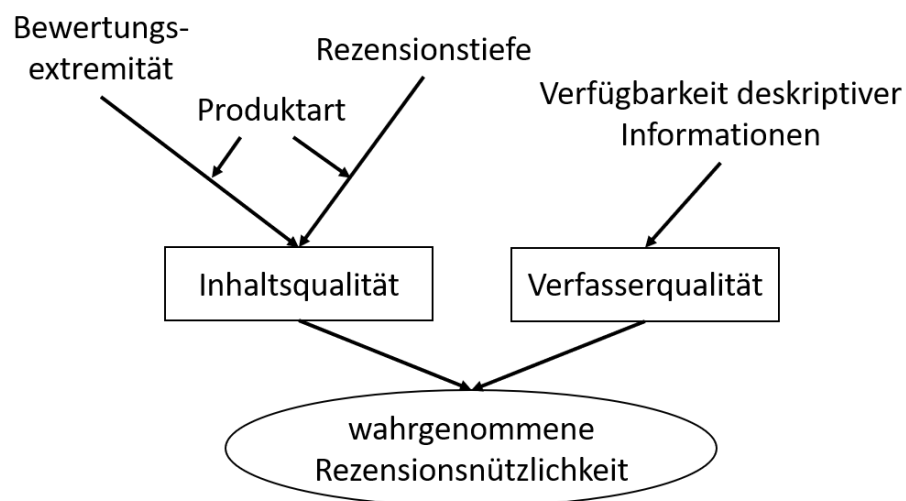
Die wahrgenommene Nützlichkeit einer Rezension wird von einigen Effekten beeinflusst, die im Folgenden näher dargestellt werden. Die Effekte werden in Elemente der Inhalts- und Verfasserqualität unterschieden (Chen et al., 2008). Unter die Faktoren der Inhaltsqualität fallen die Extremität der Bewertung und die Rezensionstiefe, gemessen an der Anzahl der Wörter. Mithilfe der Wortzahl lässt sich die langfristig wahrgenommene Nützlichkeit der Rezension großenteils vorhersagen (Ngo-Ye/Sinha, 2014). Zudem werden extreme Bewertungen als weniger hilfreich angesehen als moderate Bewertungen (Mudambi/Schuff, 2010). Dieser Effekt wird jedoch durch die Produktart (Suchgüter oder Erfahrungsgüter) moderiert. Kritischere Bewertungen sind in der Wahrnehmung nützlicher als positive Bewertungen (Chevalier/Mayzlin, 2006). Moderate Bewertungen können entweder aus einer moderaten Erfahrung mit dem Produkt oder aus guten und schlechten Erfahrungen, die sich gegenseitig auslöschen, resultieren (Kaplan, 1972; Presser/Schuman, 1980). Im letzteren Fall kann Nützlichkeit aus der Gegenüberstellung der Vor- und Nachteile des Produktes in der Textkomponente der Rezension generiert werden (Pavlou/Dimoka, 2006). Ebenfalls dem Moderationseffekt durch Produktarten unterworfen ist die Wirkung der Rezensionstiefe (Mudambi/Schuff, 2010). Unter den Faktor Verfasserqualität fallen verfügbare deskriptive Informationen über die Rezensenten. Sind diese Informationen vorhanden, wirken sie positiv auf die Evaluation der Rezension (Forman et al., 2008). Eine Übersicht über die Einflüsse auf die wahrgenommene Rezensionsnützlichkeit der Kunden ist in Abbildung 2 dargestellt. Amazon stellt ein nutzergeneriertes Maß der Nützlichkeit für jede Rezension zur Verfügung. Nutzer können zu jeder Rezension angeben, ob diese Rezension hilfreich für sie war. Die Summe der für „nützlich“ abgegebenen Stimmen wird unter jeder Rezension angezeigt. Zudem lassen sich die Rezensionen nach diesem Maß sortieren. Daraus ergibt sich ein Wettbewerb der Rezensenten um die Aufmerksamkeit der Leser, in dem sich rationale Rezensenten strategisch verhalten (Shen et al., 2015). Das Verhalten der Nutzer kann über gegebene Informationen des Nutzerverhaltens modelliert und Bewertungen für nicht gekaufte Produkte vorhergesagt werden (Fang et al., 2017). Dieses strategische Verhalten kann zu Problemen der Funktionsweise des Rezensionssystems auf Verkäuferseite führen.

3.1.2 Verkäuferseite

Ein Rezensionssystem ist anfällig für strategisches Verhalten der Rezensenten. Es bietet nicht nur die Möglichkeit, den Markt zugunsten einiger Anbieter zu manipulieren, sondern zeigt auch Fairness-Probleme in seiner eigenen Natur auf. Dabei liegt der Fokus auf vier wesentlichen Schwächen: fehlende Bewertungen, Verhaltensänderung, verfälschte Bewertungen und Identitätswechsel (Peters/Reitzenstein, 2008). Fehlende Bewertungen resultieren aus der Trittbrettfahrerproblematik im Zusammenspiel mit der freiwillig nutzbaren Möglichkeit, Rezensionen zu verfassen (Miller et al., 2005). Dementsprechend kommt es zu Informationslücken im Rezensionssystem durch fehlenden Ansporn, seine Erfahrungen mit anderen Nutzern zu teilen. Ein Anbieter hat bei einer schlechten Bewertung den Anreiz, sein Verhalten zugunsten seiner Reputation und dem damit verbundenen Absatz zu ändern (Chevalier/Mayzlin, 2006). Im Rezensionssystem behält ein solcher Anbieter allerdings seine anhaftende negative Bewertung und kann diese erst durch eine Vielzahl neuer Bewertungen ausgleichen. Dabei kommt erschwerend hinzu, dass vergangene Bewertungen als Ankerpunkt für zukünftige wirken (Wan/Nakayama, 2014). Verfälschte Bewertungen ergeben sich dabei unter anderem aus der subjektiven Wahrnehmung verschiedener Attribute. Was für den einen Kunden eine negative Eigenschaft ist, kann für einen anderen Kunden akzeptabel sein.

Geringe Identitätswechselkosten und die Möglichkeit der Erstellung von Scheinidentitäten können Produktgesamtbewertungen verzerren. Tritt der Verkäufer auch als Käufer auf, kann

Abbildung 2: Einflussfaktoren der wahrgenommenen Rezensionsnützlichkeit



das eigene Produkt mit gefälschten positiven Bewertungen (Ballot-Stuffing) besser und das Konkurrenzprodukt mithilfe gefälschter negativer Bewertungen (Bad Mouthing) schlechter dargestellt werden. Eine weitere Bedrohung durch Manipulation ergibt sich durch strategisches Verhalten der Rezensenten im Wettbewerb um Aufmerksamkeit. Hat ein Rezensent die Aufmerksamkeit auf sich gezogen und sich über die Zeit etabliert, lassen sich die Rezensionen profitmaximierend verkaufen (Coster, 2006). Im Sinne der Effizienz sollte ein Rezensionssystem robust gegen solche dominanten, manipulativen Strategien sein. Unter dieser Annahme und zusammen mit den bereits dargestellten Eigenschaften, lassen sich in Kapitel 5 Hypothesen hinsichtlich der einzelnen Elemente des Algorithmus zur Zusammenfassung von Bewertungen auf Amazon aufstellen.

3.2 Rezensionssysteme von Amazon

Amazon hat zwei Systeme in die Online-Plattform eingebettet, die Informationen bereitstellen, die von Nutzern der Plattform generiert werden. Das System „Kunden – Fragen & Antworten“ ist auf jeder Produktseite eingepflegt. Nutzer können an dieser Stelle Fragen zum Produkt stellen oder bereits gestellte Fragen nach ihrem Wissen beantworten (Amazon.de [1]). Dadurch wird Informationssuchenden die Möglichkeit geboten, sich effizient nahezu alle Informationen zu beschaffen, die sie auf Basis des bisherigen Nutzerverhaltens für ihre Kaufentscheidung benötigen. Darüber hinaus gibt es ein ausgeprägtes Kundenrezensionssystem. Nutzer können zum einen eine Bewertung auf einer Sterne-Skala von eins bis fünf abgeben und zum anderen eine Rezension per Text oder Video verfassen und anderen Nutzern zur Verfügung stellen (Amazon.de [2]). Für das Verfassen einer Rezension hat Amazon detaillierte Richtlinien veröffentlicht (Amazon.de [3]). Sie sollen dazu beitragen, das System vor Missbrauch zu schützen. Die Rezensionen sind ebenfalls auf jeder Produktseite

Abbildung 3: Aufbereitung häufig verwendeter Stichworte auf Amazon

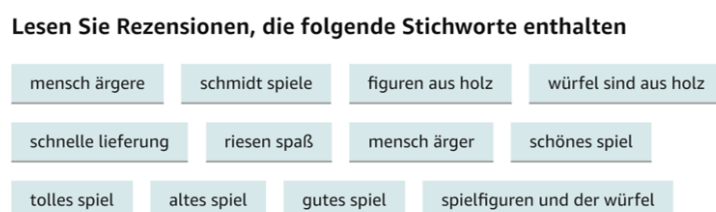


Abbildung 4: Zusammenfassende Rezensionsstatistik auf Amazon



einsehbar und können vom Seitenbesucher entweder nach Datum oder nach Nützlichkeit sortiert werden. Darüber hinaus werden die generierten Informationen aufbereitet. Von Kunden bereitgestellte Produktbilder werden separat angezeigt. Häufig auftauchende Stichworte werden ebenfalls gelistet. Abbildung 3 zeigt die Aufbereitung der häufig verwendeten Stichworte in Rezensionen. Als Beispiel wurde hier das Produkt des ersten Treffers der Suche nach „Mensch ärgere dich nicht“¹ herangezogen (Amazon.de [4]). Zusätzlich wird eine Übersicht über die abgegebenen Bewertungen, sowie ein zusammenfassendes Gesamtbewertungsergebnis dargestellt. Die grafische Aufbereitung der Zusammenfassung der abgegebenen Rezensionen zeigt Abbildung 4. Das Gesamtbewertungsergebnis wird zudem direkt unter dem Produkttitel sowohl auf der Produktseite als auch in der Produktlistung als numerischer Wert angezeigt. Außerdem wird dieser grafisch aufbereitet, mittels anteilig ausgefüllter Sterne, dargestellt. Wie im weiteren Verlauf dieser Arbeit gezeigt wird, wurde die Gesamtbewertung ursprünglich durch das arithmetische Mittel aller für das Produkt abgegebenen Bewertungen gebildet. Eine neuerliche Abweichung vom arithmetischen Mittel soll in dieser Arbeit analysiert werden.

3.3 Bedeutung einer zusammenfassenden Gesamtbewertung

Durch ein Rezensionssystem werden viele Informationen generiert, die potenzielle Kunden bei ihrer Kaufentscheidung heranziehen können. Da die Darstellung dieser unübersichtlich werden und damit den Zweck der Suchkosteneinsparung verfehlen kann, ist ein hilfreiches System ansprechend in die Benutzeroberfläche eingearbeitet und mit angemessener Informationsqualität ausgestattet (Fung/Lee, 1999). Dabei werden die Informationen häufig zugunsten der kognitiven Anstrengung der potenziellen Kunden zusammengefasst. Es wird ein einzelner vergleichbarer Wert errechnet, der Aufschluss über die Produktqualität geben soll. Duan et al. (2008) und Chen et al. (2004) zeigten jedoch, dass eine zusammenfassende Gesamtbewertung einen geringeren Einfluss auf die Produktwahrnehmung potenzieller Kunden hat als einzelne Rezensionen. Dabei hängt der Einfluss einzelner Rezensionen von ihrem Inhalt und seiner Vertrauenswürdigkeit ab (Ziegele/Weber, 2015). Duan et al. (2008) und Chen et al. (2004) erkannten lediglich, dass die Anzahl der Rezensionen als Element der zusammenfassenden Statistiken einen positiven Einfluss auf den Absatz hat. Andere Studien

¹ Mensch ärgere dich nicht: Klassisches Gesellschaftsspiel

zeigten zwar einen positiven Effekt einer hohen Gesamtbewertung auf den Absatz, doch erkannten auch, dass potenzielle Kunden auf Informationssuche einzelnen Rezensionen eine höhere Gewichtung in der Entscheidungsfindung zuordnen als der Gesamtstatistik (Chevalier/Mayzlin, 2006).

4 Algorithmen zur Bewertung und Sortierung von Elementen

In diesem Kapitel werden verschiedene Algorithmen vorgestellt und untersucht, die den Anspruch haben, auf Grundlage gegebener Informationen Elemente in eine, für ihre Nutzer hilfreiche Reihenfolge zu bringen. Dabei soll ein zusammenfassender numerischer Wert ausschlaggebend für die Einordnung der jeweiligen Elemente in die Rangfolge sein. Anschließend können Rückschlüsse auf die Kombination der Elemente in Amazons Algorithmus zur Berechnung der Produktgesamtbewertungen gezogen werden. Zunächst wird das arithmetische Mittel betrachtet, bevor auf den Histos-Algorithmus, den Sporas-Algorithmus und einen korrelationsbasierten Algorithmus eingegangen wird.

4.1 Das arithmetische Mittel

Ziel dieses Ansatzes ist es, jedem Bewertungselement einen globalen numerischen Wert R_i zuzuordnen, um anschließend anhand dieser Werte eine Rangfolge bilden zu können. Das arithmetische Mittel errechnet aus der gegebenen Anzahl an Bewertungen einen gleichgewichteten Durchschnitt. Mithilfe vergangener Bewertungen wird auf zukünftige Bewertungen geschlossen. Das arithmetische Mittel spiegelt hierfür den Erwartungswert über die Beurteilung zukünftiger Erfahrungen wider (Hübler, 1989, S. 22f.). Nach dem zentralen Grenzwertsatz nähert sich der Erwartungswert mit steigender Stichprobengröße der tatsächlichen langfristigen Erfahrungsbeurteilung an (Assenmacher, 2000, S. 176ff.). Im Folgenden ist die Formel des arithmetischen Mittels dargestellt:

$$(1) \quad R_i = \bar{r}_{ij} = \frac{1}{n_i} \sum_{j=1}^{n_i} r_{ij}$$

wobei:

R_i : Gesamtbewertung des Bewertungselementes i

n_i : Anzahl vorhandener Bewertungen über Element i

r_{ij} : Bewertung des Teilnehmers j für Element i

Das arithmetische Mittel ist dabei anfällig für Manipulationen. Verfasser können eine Bewertung abgeben, die nicht ihrer wahren Erfahrung entspricht. Diese Bewertung geht mit demselben Gewicht in die Gesamtevaluation ein wie alle anderen. Darüber hinaus wird in dieser Rohform dieses Algorithmus keine Identitätsprüfung der Teilnehmer durchgeführt, die dafür genutzt werden kann, Mehrfachbewertungen zu identifizieren und zu bereinigen. Dementsprechend kann ein nicht wohlgesonnener Nutzer manipulative Bewertungen mehrfach abgeben und auf diese Weise ihr Wirkungsgewicht erhöhen.

4.2 Der Histos-Algorithmus (nach Zacharia et al., 2000)

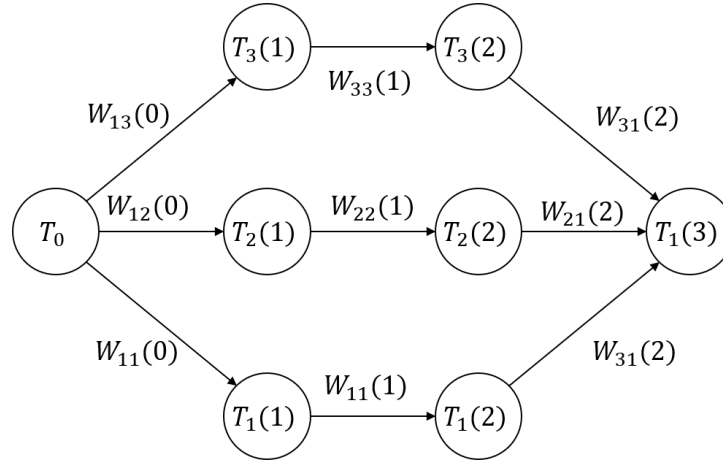
Der Histos-Algorithmus zielt darauf ab, die Vertrauenswürdigkeit aller Teilnehmer eines eng verknüpften Benutzernetzwerkes zu bewerten. Aufgrund der hohen Vernetzung und der jeweiligen abgegebenen Bewertungen der vernetzten Partner, kann für jeden Teilnehmer eine personalisierte Reputation eines anderen Teilnehmers berechnet werden. Das Netzwerk lässt sich zusammen mit den direkten Reputationen grafisch darstellen (siehe Abbildung 5). Für den Reputationswert $R_k(d)$ eines nicht direkt vernetzten Teilnehmers werden alle Bewertungen $W_{jk}(d)$ aller kürzesten Pfade zwischen den Teilnehmern in den Algorithmus einbezogen. Durch die Komplexität des Netzwerks werden Manipulationen erschwert. Ein manipulativer Teilnehmer kann nur durch eine strategisch gut geplante Platzierung im Netzwerk auf einige individuelle Bewertungen Einfluss nehmen. Im Folgenden ist die Formel des Histos-Algorithmus dargestellt:

$$(2) \quad R_k(d) = D \frac{\sum_j [R_j(d-1)W_{jk}(d)]}{\sum_j [R_j(d-1)]} \quad \forall jk, \text{ so dass } W_{jk}(d) \geq 0,5$$

wobei:

$R_k(d)$:	individuelle Reputation des k-ten Teilnehmers mit Distanz d aus Sicht von T_0
d:	Distanz zum Ausgangsnutzer
D:	höchste erreichbare Reputation
$W_{jk}(d)$:	Bewertung des j-ten Teilnehmers für k-ten Teilnehmer mit Distanz d

Abbildung 5: Reputationsnetzwerk (nach Zacharia et al., 2000)



4.3 Der Sporas-Algorithmus (nach Zacharia et al., 2000)

Ziel des Sporas-Algorithmus ist es, Nutzern eines Netzwerks einen globalen Reputationswert zuzuordnen. Dabei basiert er auf einer paarweisen Bewertung jeder Transaktion. Das bedeutet, dass sich nach jeder Transaktion die beiden Beteiligten gegenseitig aufgrund ihrer Erlebnisse bewerten. Daraufhin errechnet der Algorithmus mit den gewonnenen Informationen den neuen zusammenfassenden Gesamtreputationswert R_{t+1} . Dabei greift der Algorithmus neben den hinzugekommenen Bewertungen auf die vorherige Gesamtbewertung, die Gesamtbewertung des Bewertenden und einen zeitlichen Gewichtungsfaktor zurück. Je höher die Gesamtbewertung des Bewertenden ist, desto mehr Gewicht erhält seine abgegebene Bewertung im Reputationswert des Transaktionspartners. Die Zeitgewichtung trägt dazu bei, dass mit jeder hinzukommenden Bewertung alle vorherigen Bewertungen an Gewicht verlieren. Dadurch setzt der Algorithmus Anreize zur Qualitätsverbesserung, denn Bewertungen des aktuelleren Verhaltens fallen stärker ins Gewicht als die Evaluation vergangenen Verhaltens. Der Algorithmus funktioniert im Gegensatz zum Histos-Algorithmus auch ohne eine ausgeprägte Vernetzung der Teilnehmer. Im Folgenden sind die Formel und die Funktionen des Sporas-Algorithmus dargestellt:

$$(3) \quad R_{t+1} = R_t + \frac{1}{\theta} \phi(R_t) \cdot R_{t+1}^{other} \cdot (W_{t+1} - E(W_{t+1})) \quad \text{mit } (\theta > 1)$$

$$(4) \quad \phi(R_t) = 1 - \frac{1}{1 + e^{\frac{-(R_t - D)}{\tau}}}$$

$$(5) \quad E(W_{t+1}) = \frac{R_t}{D}$$

wobei:

R_t : alter Reputationswert

R_{t+1} : neuer Reputationswert

R_{t+1}^{other} : Reputationswert des Bewertenden

t : Anzahl der bereits erhaltenen Bewertungen

θ : Zeitgewichtung

W_{t+1} : Bewertung der neuen Transaktion

D : höchste erreichbare Reputation

τ : Beschleunigungsfaktor

Mithilfe dieses Algorithmus wird jedem Benutzer ein globaler zusammenfassender Reputationswert zugeordnet, der seine Vertrauenswürdigkeit widerspiegeln soll. Treffen Teilnehmer mehrfach aufeinander, ersetzt die neueste Bewertung die ältere Bewertung. Neueinsteiger erhalten die niedrigste Reputation von null. Die obere Grenze der Gesamtreputation ist durch D determiniert.

Es kann gezeigt werden, dass der Sporas-Algorithmus in Abhängigkeit der Parameter der Zeitgewichtung und des Beschleunigungsfaktors vor allem robust gegen Manipulationen durch eine Bad Mouthing Strategie ist (Reitzenstein/Peters, 2009). Diese Strategie bezeichnet das Platzieren von negativen Bewertungen durch Konkurrenten. Bei der Wahl der Parameter gilt es, eine Abwägung zwischen Sanktionierung und Rehabilitierung zu treffen und die Faktoren entsprechend zu wählen.

4.4 Korrelationsbasierter Bewertungsalgorithmus (nach Zhou et al., 2011)

Der korrelationsbasierte Bewertungsalgorithmus beschreibt eine iterative Berechnung eines globalen Qualitätsmaßes für angebotene Produkte. Dabei werden die Bewertungen zu jedem Produkt anhand der Korrelation aller abgegebenen Bewertungen eines Nutzers mit den zugehörigen Produktgesamtbewertungen gewichtet. Die Gleichungen des Algorithmus sind gegeben als:

$$(6) \quad Corr_j = \frac{1}{ku_j} \sum_{\alpha \in O_j} \left(\frac{r_{j\alpha} - \bar{r}_j}{\sigma_{r_j}} \right) \left(\frac{Qo_\alpha - \bar{Qo}_j}{\sigma_{Qo_j}} \right)$$

$$(7) \quad Qo_\alpha = \frac{\sum_{j \in U_\alpha} Cu_j r_{j\alpha}}{\sum_{j \in U_\alpha} Cu_j}$$

mit:

$$(8) \quad Cu_j = \begin{cases} Corr_j, & \text{falls } Corr_j \geq 0, \\ 0, & \text{falls } Corr_j < 0. \end{cases}$$

wobei:

$Corr_j$: Korrelation der Bewertungen des j-ten Teilnehmers mit den jeweiligen Produktgesamtbewertungen

ku_j : Anzahl der abgegebenen Bewertungen des j-ten Teilnehmers

$r_{j\alpha}$: abgegebene Bewertung des j-ten Teilnehmers für Objekt α

\bar{r}_j : Erwartungswert der Bewertung des j-ten Teilnehmers

\bar{Qo}_j : Erwartungswert der geschätzten Gesamtbewertungen der vom j-ten Teilnehmer bewerteten Produkte

Qo_α : aggregierte geschätzte Gesamtbewertung des Objektes α

σ_v : Standardabweichung des Vektors v

O_j : Menge der Objekte, die vom j-ten Teilnehmer bewertet wurden

U_α : Menge der Nutzer, die Objekt α bewertet haben

Cu_j : Reputation (Gewichtung) des j-ten Teilnehmers

Mithilfe des Korrelationskoeffizienten wird die Plausibilität der abgegebenen Bewertungen quantifiziert und in den Bewertungsalgorithmus aufgenommen. Da sich jede von einem Nutzer hinzukommende Bewertung auf die Gesamtbewertung auswirkt und dadurch Einfluss auf die Korrelation (Gewichtung) hat, ist ein iterativer Prozess notwendig. Der Prozess wird mit jeder einzelnen neuen Bewertung für die Gesamtbewertung aller im Rezensionssystem enthaltenen Objekte angeregt. Er stoppt, sobald die Änderungen der Gesamtbewertungen äußerst gering sind. Mit wachsender Objektzahl und steigender Nutzeraktivität ist damit ein großer Rechenaufwand verbunden. Der Algorithmus erweist sich jedoch als robust gegenüber Manipulationen, weil sich manipulative Handlungen sofort in ihrer Gewichtung der Bewertung niederschlagen.

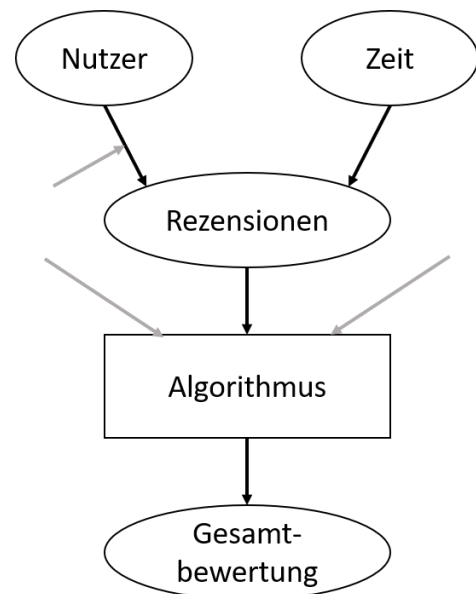
5 Konzeptioneller Rahmen

Auf der Internetplattform Amazon gibt es ein ausgeprägtes Rezensionssystem zur Bewertung im Angebot stehender Produkte. Innerhalb dieses Rezensionssystems wird eine Gesamtbewertung errechnet, die den Plattformbesuchern zu jedem Produkt in der zusammenfassenden Gesamtstatistik des Rezensionssystems zur Ansicht gestellt wird. Dieser Wert wird numerisch und grafisch aufbereitet dargestellt. Der Algorithmus zur Errechnung dieser Gesamtbewertung ist Kernuntersuchungsgegenstand dieser Bachelorarbeit. Es gilt zu untersuchen, welche Faktoren Einfluss auf diesen Wert haben. In den vergangenen Kapiteln wurden dazu die Vertrauensprobleme im Online-Handel, die Rolle und Eigenschaften von Rezensionssystemen erläutert, sowie verschiedene Algorithmen zur Errechnung einer Gesamtbewertung in Rezensionssystemen dargestellt. Ziel ist es nun, die gewonnenen Erkenntnisse zu bündeln und Hypothesen über die einzelnen Elemente und ihre Verknüpfungen im Bewertungsalgorithmus von Amazon aufzustellen.

Als Online-Plattform steht Amazon vor dem Vertrauensproblem von Transaktionen im Internet. Es ist davon auszugehen, dass das Rezensionssystem von Amazon als Instrument zur Vertrauensbildung eingesetzt wird. Die Gesamtbewertung, ein Element dieses Systems, wird dies dementsprechend ebenfalls zum Ziel haben. Sie hat den Anspruch die von anderen Kunden wahrgenommene Qualität des Produktes widerzuspiegeln, um andere potenzielle Kunden bei der Kaufentscheidung zu unterstützen. Dadurch können die vom Anbieter zur Verfügung gestellten Bilder und Beschreibungen des Produktes durch Informationen einer dritten Partei ergänzt und damit die von potenziellen Kunden wahrgenommene mögliche Qualitätsvarianz eingedämmt werden. Um die von den Kunden wahrgenommene Qualität des Produktes in einem einzigen Wert zusammenzufassen, müssen alle abgegebenen Bewertungen so kombiniert werden, dass sich ein repräsentativer und interpretierbarer Wert ergibt. Um die Ausmaße dieser Arbeit im vertretbaren Rahmen zu halten, werden die untersuchten Datenquellen des Algorithmus vorab eingeschränkt. Diese Arbeit untersucht ausschließlich mögliche Einflussfaktoren, die auf Daten basieren, die durch das Rezensionssystem selbst generiert werden. Es ist dabei jedoch nicht vollkommen auszuschließen, dass das System Informationen von außerhalb bezieht. Diese externen Einflussfaktoren gilt es, in weiterführenden Forschungen zu untersuchen. Die einzige selbstgenerierte Informationsquelle des Systems sind die nutzerbasierten Rezensionen. Sämtliche rezensionsgebundenen Daten

können als Element des Gesamtbewertungsalgorithmus in Betracht bezogen werden. Das in dieser Arbeit unterstellte Grundkonzept des Algorithmus ist in Abbildung 6 dargestellt. Dunkle Pfeile stellen die in dieser Arbeit betrachteten und helle mögliche, aber in dieser Arbeit außer Acht gelassene Einflüsse dar. Wie in Kapitel 4 erläutert, ist eine interpretierbare Zusammenfassung durch das arithmetische Mittel aller abgegebenen Bewertungen möglich. Diese Variante hat die Eigenschaft, dass die Gesamtbewertung auf derselben Skala interpretierbar ist wie die abgegebenen Einzelbewertungen. Der Wert ist sowohl nach oben als auch nach unten begrenzt und innerhalb dieser Grenzen lassen sich die Produkte hinsichtlich ihrer Bewertung anordnen. Nutzer von Amazon können Bewertungen auf einer Sterneskala von eins (sehr schlecht) bis fünf (sehr gut) abgeben. Auch der Gesamtwert rangiert innerhalb dieser Grenzen. Es kann also vermutet werden, dass Amazons Gesamtbewertungsalgorithmus auf dem arithmetischen Mittel aller abgegebenen Bewertungen basiert. Die Hypothese wird in den folgenden Kapiteln überprüft. Dieser Algorithmus ist ausschließlich dann repräsentativ für die tatsächliche Produktqualität, wenn manipulativen Handlungen der Bewertenden wirkungslos sind. Im Sinne der Robustheit gegen solche Angriffe ist zu vermuten, dass Amazons Algorithmus durch weitere Elemente ergänzt wurde.

Abbildung 6: Konzeptioneller Rahmen der Arbeit



Für das Funktionieren eines Rezensionssystems und der Zusammenfassung seiner Inhalte ist es wichtig, dass potenzielle Kunden wissen, dass sie sich auf die angegebenen Werte verlassen können. Ist das nicht der Fall, verfehlt das System sein Ziel des Vertrauensaufbaus. In Kapitel 3 wurde dargestellt, dass potenzielle Kunden die vorhandenen Rezensionen unterschiedlich evaluieren und in ihrem Entscheidungsprozess entsprechend gewichten. Wenn Informationssuchende diese Evaluation vornehmen, ist es plausibel, eben diese Gewichtung im Gesamtbewertungsalgorithmus zu berücksichtigen, um die Zusammenfassung aus Sicht der Kunden repräsentativ zu gestalten. Unter der Annahme, dass Kunden manipulative Bewertungen weniger stark gewichten, würde eine Hinzunahme dieser Gewichte in den Algorithmus die Gesamtbewertung robuster gegen Manipulationen machen.

Neben der Kundenseite gilt es auch den Einfluss der Gesamtbewertung auf die Verkäuferseite zu beachten. Wird jede Rezension gleich stark im Algorithmus gewichtet, gibt es für Anbieter keinen Anreiz zur Veränderung der Produktqualität. Neue Rezensionen nach veränderter Qualität hätten relativ wenig Einfluss auf die Gesamtbewertung, sodass die aktuelle Qualität des Verkäufers nicht repräsentativ abgebildet werden würde. Es lässt sich daher vermuten, dass eine Komponente im Amazon-Algorithmus existiert, die einen Anreiz für Anbieter setzt, ihre Qualität zu verbessern. Denkbar ist eine zeitliche Gewichtung, die vergleichbar mit der entsprechenden Komponente im Sporas-Algorithmus ist. Der Sporas-Algorithmus ist effektiv bei losen Verbindungen zwischen Transaktionspartnern, wie es bei den Kunden und Verkäufern auf einer Plattform wie Amazon der Fall ist. Mit einer solchen Komponente würden Bewertungen vergangener Qualität über die Zeit weniger Gewicht erhalten, das heißt, zunehmend in Vergessenheit des Systems geraten.

Es wird angenommen, dass Amazon daran interessiert ist, ein möglichst hilfreiches Gesamtmaß zur Beurteilung der Produktqualität bereitzustellen. Über das arithmetische Mittel hinaus liefert die erörterte vorhandene Literatur starke Evidenz zu der Hypothese, dass eine zeitliche Gewichtung der Rezensionen in Amazons Algorithmus stattfindet. Deswegen ist es Ziel dieser Arbeit, den Gesamtbewertungsalgorithmus von Amazon auf eine zeitliche Komponente in der Gewichtung der einzelnen Rezensionen zu untersuchen.

6 Design der empirischen Studie

Beim Rekonstruieren von Algorithmen ist es vorteilhaft, zunächst den Grundbaustein des Algorithmus zu identifizieren. In Kapitel 5 konnte die Hypothese entwickelt werden, dass es sich im Fall des Amazon-Algorithmus um das arithmetische Mittel aller zum Produkt abgegebenen Bewertungen handelt. Um dies zu prüfen, werden für verschiedene Produkte die Anzahl aller Bewertungen, aufgeschlüsselt nach ihrer Bewertungshöhe, benötigt. Mithilfe der jeweiligen Anzahl der abgegebenen Bewertungen kann das arithmetische Mittel errechnet werden. Der errechnete Wert für jedes Produkt kann anschließend mit der zugehörigen tatsächlichen Gesamtproduktbewertung verglichen werden. Dazu werden die Daten über diese Referenzwerte benötigt. Ist eine Abweichung zwischen den beiden Werten zu sehen, kann das arithmetische Mittel nicht vollständig die Produktgesamtbewertung erklären. Das bedeutet, dass zumindest das arithmetische Mittel allein nicht dem Amazon-Algorithmus

entsprechen kann. Stimmen die beiden verglichenen Werte überein, ist davon auszugehen, dass der Algorithmus die Gesamtbewertung durch das arithmetische Mittel aller Bewertungen für das Produkt bildet.

Falls eine Abweichung zwischen errechnetem arithmetischem Mittel und tatsächlichem Referenzwert gefunden wird, gilt es, den Algorithmus auf seine weiteren Elemente zu untersuchen. Dieser Arbeit vorangegangene Studien liefern verschiedene Ansätze zur Rekonstruktion von Algorithmen. Beel/Gipp (2009) extrahierten einflussreiche Elemente im Rangfolgenalgorithmus von Google Scholar. Hierbei wurden die Zusammenhänge visuell über grafische Darstellungen der Daten ermittelt. Chen et al. (2016) untersuchten algorithmische Preissetzung verschiedener Anbieter auf Amazon. Zur Identifikation einflussreicher Elemente wurde maschinelles Lernen eingesetzt. In dieser Arbeit geht es um die Untersuchung des Algorithmus auf eine zeitliche Gewichtung. Dabei kann im Vorfeld keine Annahme darüber getroffen werden, wie diese zeitliche Komponente mit anderen Elementen des Algorithmus verknüpft ist. Aufgrund dieser Ausgangslage wird zur Untersuchung dieses Faktors überwachtes maschinelles Lernen eingesetzt. Diese Form des maschinellen Lernens bezeichnet das automatische Identifizieren von Zusammenhängen in Datensätzen (Kelleher et al., 2015, S. 3). Mithilfe einer abhängigen und einer oder mehrerer unabhängigen Variablen wird ein Modell trainiert, das auf Grundlage des Gelernten ein Modell formt, das mit gegebenem Input Vorhersagen treffen kann. Für das Trainieren des Modells gibt es verschiedene Ansätze, die sich in ihren Anforderungen an die Variablen und Vorgehensweisen unterscheiden. In dieser Arbeit werden verschiedene Ansätze des maschinellen Lernens herangezogen, die anhand ihrer Vorhersagegenauigkeit evaluiert und verglichen werden. Das genaueste Modell kann daraufhin identifiziert werden und hinsichtlich der Wichtigkeit der unabhängigen Variablen genauer betrachtet werden.

Während der Trainingsphase der Probelaufe wird mit einer zehnfachen Kreuzvalidierung gearbeitet, um eine Überanpassung der Modelle zu vermeiden. Dabei wird der Datensatz in zehn gleich große Partitionen unterteilt. Neun der Partitionen werden zum Trainieren der Modelle verwendet und der übrige Teil zur Validierung des Modells (Kelleher et al., 2015, S. 408ff.). Dieser Vorgang wird zehnfach wiederholt. Die Wahl der zehnfachen Variante wurde in dieser Arbeit auf Grundlage von in der Literatur üblichen Werten getroffen (Knox,

2018, S. 133ff.). Mittels dieses Verfahrens können Genauigkeitswerte ermittelt werden, die später zur Evaluation der Modelle dienen.

Zur Evaluation der Modelle wird vorrangig das Maß der Wurzel der mittleren Fehlerquadratsumme (RMSE) verwendet. Das Maß wird mithilfe der Teststichprobe der Kreuzvalidierung bestimmt. Je kleiner dieser Wert ist, desto präziser ist die Vorhersage des Modells. Daher wird nach dem Trainieren der einzelnen Modelle dasjenige ausgewählt, welches die niedrigste RMSE aufweist (Burger, 2018, S. 87f.).

$$(9) \quad RMSE = \sqrt{\frac{\sum_i^N (\hat{y} - y)^2}{N}}$$

Für maschinelles Lernen wird ein umfangreicher Datensatz benötigt, um das Modell ausreichend trainieren zu können. Da die Gesamtbewertung der Amazon-Produkte untersucht wird, werden diese Daten als abhängige Variable benötigt. Darüber hinaus sind Daten über eine Aufschlüsselung der Rezensionen nach Produkt und Bewertungshöhe nötig. Um die zeitliche Komponente zu untersuchen, bedarf es darüber hinaus Informationen über die Zeitpunkte der abgegebenen Bewertungen und der Aufzeichnung der Gesamtbewertung. Zusätzlich ist es erforderlich, dass die Aufzeichnung über einen ausreichend langen Zeitraum stattfindet. Die Änderung der Gesamtbewertung soll damit durch zeitliche Variablen und die Höhe der abgegebenen Bewertungen erklärt werden.

Im ausgewählten finalen Modell werden die ermittelten optimalen Parameter verwendet, um die Maschine zu trainieren. Für das Training wird der Datensatz in zwei ungleich große Partitionen zufällig unterteilt. Die Trainingsdaten für das Modell machen 80% des gesamten verwendeten Datensatzes aus. Die übrigen 20% der Daten werden anschließend zum Überprüfen der Güte des Modells verwendet. Die verschiedenen Methoden werden mithilfe von R programmiert und durchgeführt. Dazu werden folgende Pakete verwendet: caret, kernlab, elasticnet, randomForest, gbm und plyr. Nachfolgend werden die verwendeten Modelltypen näher erläutert.

6.1 Stützvektormaschinen (SVM)

Der SVM-Ansatz ist ein Ansatz des fehlerbasierten Lernens. Die Datenpunkte werden anhand der Variablen in einen mathematischen Raum eingeordnet. Anschließend errechnet die

Maschine eine möglichst optimale trennende Hyperebene zwischen den unterschiedlichen Ausprägungen der abhängigen Variablen. Eine optimale Hyperebene ist dabei dadurch gekennzeichnet, dass die nächsten Datenpunkte möglichst weit von der Hyperebene entfernt sind. Dazu wird eine möglichst große Hülle um die Hyperebene gebildet, in der sich keine Datenpunkte befinden (Knox, 2018, S. 82ff.). Dabei spielen nur die nächsten Datenpunkte zu der Hyperebene eine Rolle. Diese bilden den Stützvektor für die Hülle der Hyperebene. Weiter entfernte Datenpunkte haben keinen Einfluss auf die Position der Hyperebene, da sie außerhalb des kritischen Bereichs liegen und ohnehin hinsichtlich ihrer Ausprägung der AV durch die Hyperebene eindeutig zugeordnet werden können. Um die Hyperebene zu bestimmen, muss ein gewisser Fehlerrahmen zugelassen werden, da es sonst möglich ist, dass keine solche Hyperebene existiert (Alpaydin, 2008, S. 235). Im Ursprung sieht das SVM-Modell eine lineare Hyperebene vor, die die Daten binär klassifiziert. Die Erweiterung durch ein polynomisches Kernel lässt jedoch auch Hyperebenen höheren Grades zu. Des Weiteren ist es mittels eines paarweisen Vergleiches möglich, eine Trennung des Datensatzes in mehrere Klassen vorzunehmen, sowie Regressionen durchzuführen (Mohanty et al., 2013, S. 254; Alpaydin, 2008, S. 239f.). Der zugelassene Fehler ist bei dieser Variation der SVM wichtig, um ein überangepasstes Modell zu vermeiden. Dabei gibt es zwei Arten der zugelassenen Fehler: zum einen diejenigen Datenpunkte, die hinsichtlich der abhängigen Variable auf der falschen Seite der Trennung sind, und zum anderen diejenigen Datenpunkte, die hinsichtlich ihrer abhängigen Variable zwar auf der richtigen Seite sind, jedoch innerhalb der Hülle liegen (Alpaydin, 2008, S. 235f.). Bei der Entwicklung eines Modells aus einem SVM-Ansatz mit polynomischem Kernel müssen im Vorfeld folgende optimale Parameter gefunden werden: der Grad des Kernels und die zugelassenen Fehler (Scale, C).

6.2 Elastische Netze

Eine weitere Form des maschinellen Lernens bietet die Regression durch Elastische Netze. Dabei handelt es sich um eine Zusammenführung der Grundgedanken der Ridge-Regression und der Lasso-Regression. Die Ridge-Regression basiert auf der Technik der OLS-Regression (Regression mithilfe der Methode der kleinsten Quadrate). Bei einer OLS-Regression wird die Summe der quadrierten Residuen (SSR) minimiert. Im Konzept des maschinellen Lernens bringt dieser Minimierungsansatz ein wesentliches Problem mit sich. Es besteht die Gefahr, die Regressionslinie auf den Trainingsdatensatz überanzupassen und

damit keine akkuraten Vorhersagen treffen zu können. Es ergibt sich somit ein Abtausch zwischen Varianz und Verzerrung (Hastie et al., 2009, S. 37f.). Bei der OLS-Regression liegt der Fokus auf der Minimierung der Verzerrung zulasten der Varianz. Mit Varianz wird in diesem Zusammenhang die Abweichung zwischen der Anpassung der Regressionslinie an die Trainingsdaten und ihrer Anpassung an die Testdaten bezeichnet. Um das Problem der Überanpassung an die Trainingsdaten zu lösen, wird bei der Ridge-Regression eine sogenannte Bestrafung in das Minimierungsproblem eingebaut. Dementsprechend wird die Summe aus der SSR und der Strafe minimiert. Dabei besteht die Strafe aus der Summe der quadrierten OLS-Steigungskoeffizienten und einem Faktor λ_1 . Daraus ergibt sich folgendes Minimierungsproblem (Knox, 2018, S. 28f.):

$$(10) \quad \min_{\beta_l} [SSR + \lambda_1 \sum_{l=1}^m \beta_l^2]$$

wobei:

β_l	l-ter Steigungsparameter
SSR	Summe der quadrierten Residuen
λ_1	exogener Faktor
m	Anzahl der Steigungsfaktoren im Modell

Durch die additive Verknüpfung des bestrafenden Terms mit der SSR findet keine ideale Anpassung der Regressionslinie an die Trainingsdaten im Sinne einer OLS-Regression statt. Durch das Zulassen einer Abweichung von dieser OLS-Anpassung kann der Abtausch zwischen Verzerrung und Varianz des Modells optimiert werden. (Knox, 2018, S. 97ff.). Dabei kann über λ_1 die zugelassene Variabilität der Parameter bestimmt werden (Alpaydin, 2008, S. 279).

Die Lasso-Regression bedient sich derselben Grundidee wie die Ridge-Regression (Burger, 2018, S. 78ff.). Während im Minimierungsproblem der Ridge-Regression die Steigungsparameter quadriert werden, arbeitet die Lasso-Regression mit ihren Beträgen. Daraus ergibt sich folgendes Minimierungsproblem:

$$(11) \quad \min_{\beta_l} [SSR + \lambda_2 \sum_{l=1}^m |\beta_l|]$$

wobei:

λ_2 exogener Faktor

Das führt dazu, dass die Steigungsparameter nicht nur, wie bei der Ridge-Regression, in Richtung null tendieren können, sondern auch diesen Wert annehmen können. Dadurch lässt dieser Ansatz die Identifikation nicht beeinflussender Variablen zu.

Der Ansatz des Elastischen Netzes ergänzt die Gedanken der Ridge-Regression mit denen der Lasso-Regression und profitiert von den jeweiligen Vorteilen der beiden Techniken. Dabei bildet sich ein kombiniertes Minimierungsproblem:

$$(12) \quad \min_{\beta_l} [(SSR + \lambda_2 \sum_{l=1}^m |\beta_l|) + (SSR + \lambda_1 \sum_{l=1}^m \beta_l^2)]$$

Bei der Entwicklung eines Modells auf Grundlage des Ansatzes des Elastischen Netzes müssen dementsprechend zwei optimale Parameter vorab bestimmt werden: λ_1 und λ_2 . Diese Parameter werden im Skript und in der Analyse mit „lambda“ und „fraction“ bezeichnet.

6.3 Zufallswälder

Zufallswälder kombinieren den Ansatz der Entscheidungsbäume mit der Technik des bagging². Entscheidungsbäume bestehen aus mehreren Ebenen, wobei für jede unabhängige Variable eine Ebene gebildet wird. Von jedem Entscheidungsknoten der Ebene gehen Äste in der Anzahl der Merkmalsausprägungen aus. Die Gesamtheit der Entscheidungsknoten und deren Äste ergeben einen Entscheidungsbaum. Jedem Entscheidungsknoten werden Wahrscheinlichkeiten für die einzelnen Äste zugeordnet, auf deren Grundlage in Verbindung mit dem Eingabewert eine Entscheidung getroffen wird (Alpaydin, 2008, S. 185ff.). Diese Wahrscheinlichkeiten werden mithilfe des Eingabedatensatzes trainiert. Durch neue Eingabewerte für die unabhängigen Variablen trifft der Entscheidungsbaum eine, auf den erlernten Wahrscheinlichkeiten basierende, Vorhersage über die Ausprägung der abhängigen Variablen. Um die Güte der Vorhersage zu verbessern, werden im Ansatz der Zufallswälder mehrere Entscheidungsbäume gepflanzt und eine Entscheidung anhand der Gesamtheit der Vorhersagen aller Entscheidungsbäume getroffen (Alpaydin, 2008, S. 384f.). Dabei erhalten alle

² bagging (Bootstrap, Aggregating) beschreibt ein Verfahren, in dem über mehrere Bootstrap-Stichproben aggregiert wird (Alpaydin, 2008, S. 385)

Entscheidungsbäume das gleiche Gewicht im Einfluss auf die finale Entscheidung. Mit der Anzahl der Entscheidungsbäume wird die Varianz der Gesamtentscheidung, unter der Annahme, dass die Bäume untereinander wenig korreliert sind, geringer (Hastie et al., 2009, S. 587ff.). Ein Vorteil des Zufallswälder-Ansatzes gegenüber anderen Ansätzen des maschinellen Lernens ist seine Robustheit gegenüber Überanpassung, beispielsweise durch zu viele gepflanzte Bäume (Knox, 2018, S.116f.). Zur Bildung der Entscheidungsbäume wird eine festgelegte Anzahl an Bootstrap-Stichproben gezogen, die jeweils die Grundlage für die Entwicklung eines Baumes bilden. Diese Unterstichproben haben dabei die Eigenschaft, sich untereinander ähnlich, aber nicht gleich zu sein, da sie derselben Überstichprobe entstammen. Die große Ähnlichkeit kann für eine starke Korrelation sorgen. Deswegen ist es üblich, dass nicht alle unabhängigen Variablen in jeder Bootstrap-Stichprobe vorkommen. Dadurch wird eine Korrelation zwischen den Bäumen des entstehenden Waldes vermieden. Aus diesem Grund muss zunächst die Anzahl der Variablen pro Baum bestimmt werden. Bei der Entwicklung eines Modells aus einem Zufallswald-Ansatz des maschinellen Lernens müssen vorab folgende optimale Parameter festgelegt werden: die Anzahl der zu pflanzenden Bäume und die Anzahl der unabhängigen Variablen in jeder Bootstrap-Stichprobe (mtry).

6.4 Gradient Boosting Maschine (GBM)

Der Ansatz der Gradient Boosting Maschine basiert wie die Zufallswälder auf dem Konzept der Entscheidungsbäume. Ziel ist hierbei jedoch nicht die Pflanzung mehrerer voneinander unabhängiger Bäume, sondern die iterative Entwicklung eines additiven Gesamtmodells. Bei jeder Iteration werden dazu neue Parameter errechnet, die die Verlustfunktion unter den Gegebenheiten der Ergebnisse der vorherigen Iteration minimieren (Hastie et al., 2009, S. 359). Die Parameter werden bei jeder Iteration additiv mit dem vorherigen Modell verknüpft, sodass der Verlust mit jeder Iteration geringer wird. Damit ist die Anzahl der Iterationen ausschlaggebend für die Anpassung des Modells. Wird diese Anzahl zu hoch festgesetzt, kann es zur Überanpassung des Modells kommen (Hastie et al., 2009, S. 364). Auf Grundlage der Modellfunktion nach der neuesten Iteration wird die Richtung des negativen Gradienten ermittelt. Daraufhin wird ein Entscheidungsbaum entwickelt, der die festgelegte Verlustfunktion minimiert. Hierbei kann eine Lernrate festgesetzt werden, die dafür sorgt, dass das Modell mit jeder Iteration nur gedämpft in Richtung des negativen Gradienten angepasst wird. Dabei ist zu beachten, dass es eine Wechselwirkung zwischen der Anzahl an

Iterationen und der Lernrate gibt. Wird die Lernrate gering gewählt, werden die Testergebnisse des Modells hinsichtlich des Fehlers verbessert, jedoch ist dafür eine größere Anzahl an Iterationen nötig (Friedman, 2001). Anschließend werden die Parameter des Entscheidungsbaums zur Gradient Boosting Maschine additiv hinzugefügt (Hastie et al., 2009, S.359ff.). Für die Modellentwicklung ist es nötig, die Anzahl der Ebenen der Entscheidungsbäume festzulegen. Daher werden verschiedene Tiefen der Bäume in Probeläufen überprüft und der beste Wert festgelegt. Dementsprechend sind für den Ansatz der Gradient Boosting Maschine folgende optimale Parameter festzulegen: die Anzahl der Iterationen (Anzahl der Bäume), die Lernrate und die Tiefe der Bäume.

7 Datensatz

Dieser Bachelorarbeit liegt ein umfangreicher Datensatz zugrunde, der von Trutz Fries³ (Amalytix) zur Bearbeitung des Themas im Rahmen dieser Bachelorarbeit zur Verfügung gestellt wurde. Amalytix ist ein Unternehmen, das sich darauf spezialisiert hat, Amazon-Produkte seiner Kunden in ihrem Auftrag automatisiert zu überwachen (Amalytix.com [2]). Dadurch generiert das Unternehmen die Masse an Daten, die für diese Bachelorarbeit zur Verfügung stehen. Die Gesamtzahl der Datenpunkte beläuft sich auf etwa 48,7 Mio. Die Daten beinhalten tägliche Aufzeichnungen über Rezensionsdaten von 346.440 verschiedenen Produkten, die auf der Internetseite von Amazon angeboten werden. Der Datensatz enthält ausschließlich Informationen über Produkte, die auf dem deutschen Amazon-Markt angeboten wurden. Dabei stammen die Produkte aus verschiedenen Kategorien. Die Produkte lassen sich anhand ihrer Amazon-Standard-Identifikationsnummer (ASIN) unterscheiden und im Datensatz eindeutig identifizieren. Aufgezeichnet wurden im Gesamtzeitraum vom 21.07.2016 bis 29.01.2019 Informationen über die absolute Anzahl der vorhandenen Rezensionen zum entsprechenden Produkt, die absolute Anzahl der Rezensionen pro Bewertung (eins bis drei, sowie fünf) und die Gesamtbewertung, die auf der Amazon-Produktseite angezeigt wird. Dabei liegen die Datenpunkte für jedes Produkt zwar immer zeitlich zusammenhängend, aber nicht für den gesamten oben genannten Zeitraum vor. Bei späteren Analysen wurden die Zeiträume der Aufzeichnungen für jedes Produkt beachtet und auf

³ Trutz Fries: Inhaber von Amalytix (Amalytix.com [1])

Relevanz geprüft. Um den Datensatz an die verwendeten Analysemethoden anzupassen, mussten einige Änderungen und Ergänzungen vorgenommen werden.

7.1 Ursprünglicher Datensatz

7.1.1 Strukturelle Veränderung des Datensatzes

Zunächst konnten einige fehlende Werte in der Aufzeichnung der Produktgesamtbewertungen und der Anzahl der Rezensionen identifiziert werden. Diese Datenspalten werden im weiteren Verlauf dieser Arbeit wesentliche Rollen spielen, da sie entweder eng mit der abhängigen Variablen oder mit allen unabhängigen Variablen verknüpft sind. Diese fehlenden Werte konnten nicht durch gängige Methoden sinnvoll ersetzt werden. Daher wurden entsprechend fast 21 Mio. Datenpunkte für die weitere Bearbeitung entfernt. Um eine gute Analysequalität garantieren zu können, ist es nötig, dass eine Mindestanzahl an Daten für jedes Produkt vorhanden ist. Da die Analyse der zeitlichen Entwicklung der Gesamtbewertungen im Vordergrund dieser Arbeit steht, werden nur Produkte in die Analyse miteinbezogen, deren zusammenhängender Aufzeichnungszeitraum länger als ein Jahr ist. Dadurch kann gewährleistet werden, dass sich eine zeitliche Gewichtung in der Veränderung der Gesamtbewertungen niederschlagen würde. Um spätere Anforderungen an die Rechenleistung zu minimieren, wird der Datensatz auf Datenpunkte reduziert, die eine Veränderung hinsichtlich der Gesamtbewertung oder Rezensionsanzahl zu ihrem zeitlichen Vorgänger aufweisen. Datenpunkte, die keine zeitliche Veränderung aufweisen, haben keinen Einfluss auf das Ergebnis der später angewendeten Analyseverfahren.

7.1.2 Ergänzungen im Datensatz

Zunächst wird der ursprüngliche Datensatz durch das arithmetische Mittel der Bewertungen der abgegebenen Rezensionen ergänzt. Die Daten enthalten lediglich Informationen über die jeweilige Anzahl an Rezensionen für die Bewertungen eins bis drei, sowie fünf. Daher wurden Daten über die Anzahl der Bewertungen der Höhe vier ergänzt. Diese konnten mithilfe der Differenz zwischen der Gesamtanzahl der Rezensionen und der Summe aller vorhandenen, nach Bewertungshöhe aufgeschlüsselten, jeweiligen Anzahl der Rezensionen ermittelt werden. Darüber hinaus wurde der Datensatz durch fünf Spalten ergänzt, die für jede Bewertungshöhe in jeder Datenzeile die Differenz der Anzahl der Rezensionen zum zeitlichen Vorgänger angeben.

7.2 Datensatz der minimalen und maximalen Bewertungen

Aus dem ursprünglichen Datensatz konnte für jedes Produkt die jeweils höchste und niedrigste Gesamtbewertung im Aufzeichnungszeitraum ermittelt werden. Daraufhin wurde ein neuer Datensatz generiert, der Informationen über die ASIN, die zugehörige minimale und maximale Gesamtbewertung im Aufzeichnungszeitraum, das Datum, an dem die jeweiligen minimalen und maximalen Gesamtbewertungen erreicht wurden, und die jeweilige Gesamtanzahl der Rezensionen, die an den beiden Daten zu beobachten war, enthält.

7.2.1 Strukturelle Veränderungen des Datensatzes

Entsprechen sich die minimale und maximale Gesamtbewertung, liegt im gesamten Aufzeichnungszeitraum keine Veränderung vor. Sollte dies der Fall sein, können aus dieser Datenaufzeichnung keine Informationen gewonnen werden, sodass diese Produkte aus dem Datensatz entfernt werden.

7.2.2 Ergänzungen im Datensatz

Für spätere Berechnungen wird dem Datensatz die Differenz in Tagen zwischen dem Datum, an dem die höchste Bewertung erreicht wurde und dem Datum, an dem die niedrigste Bewertung erzielt wurde, hinzugefügt. Die Differenz zwischen höchster und niedrigster Gesamtbewertung im Beobachtungszeitraum wird ebenfalls ergänzt. Zu jedem Datenpunkt im Datensatz der minimalen und maximalen Bewertungen werden nun die Änderungen der einzelnen Rezensionszahlen pro Bewertungshöhe aus dem ursprünglichen Datensatz ermittelt. Dabei werden die Änderungen jeweils nach Zuwachs und Abgang getrennt summiert, um nachvollziehen zu können, wie viele neue Rezensionen tatsächlich im Beobachtungszeitraum hinzugekommen sind. Der Datensatz wurde zusätzlich mit dem Durchschnittsalter der Rezensionen pro Bewertungshöhe ergänzt.

7.3 Finaler Datensatz

Aus dem Datensatz der minimalen und maximalen Bewertungen kann schlussendlich der finale Datensatz gebildet werden. Dazu werden die Differenz der höchsten und niedrigsten Bewertung eines Produktes, der Zeitraum in Tagen zwischen der höchsten und niedrigsten Gesamtbewertung des Produktes, die Gesamtanzahl der Rezensionen zum Zeitpunkt der höchsten Bewertung eines Produktes im Aufzeichnungszeitraum, die Änderung der Anzahl der Rezensionen, aufgeschlüsselt nach ihrer Bewertungshöhe und getrennt nach Zu- und

Abgang, sowie das Durchschnittsalter der Rezensionen, ebenfalls aufgeschlüsselt nach ihrer Bewertungshöhe und getrennt nach Zu- und Abgang, verwendet. Abschließend werden noch einige Variablen hinsichtlich ihrer Interpretierbarkeit angepasst. Im Folgenden werden die dafür nötigen Veränderungen verdeutlicht. Das Durchschnittsalter der Rezensionen wird pro Bewertungshöhe hinsichtlich der Intervalllänge relativiert. Die Intervalllänge beschreibt die zeitliche Differenz zwischen höchster und geringster Gesamtbewertung eines Produktes. Um den Einfluss der im Zeitintervall hinzukommenden Rezensionen abzubilden, wird die Anzahl dieser Rezensionen pro Bewertungshöhe anhand der Gesamtzahl der Rezensionen relativiert. Zusätzlich werden sie dazu mit ihrer jeweiligen Bewertungshöhe multipliziert. Der Datensatz, der im weiteren Verlauf in der Analyse verwendet wird, hat eine Größe von $n = 6.238$. In Tabelle 1 werden deskriptive Statistiken zum Veranschaulichen der Datenstruktur dargestellt. Die unabhängigen Variablen werden dabei am Beispiel der Daten für die mittlere Bewertung veranschaulicht. Die Datenstrukturen aller Variablen sind im Anhang (Tabelle 4) zu finden. Sie weisen eine ähnliche Struktur auf wie das hier angeführte Beispiel. Die später als abhängig verwendete Variable „ratingdifferenz“ bezeichnet die Differenz zwischen der höchsten und geringsten Gesamtbewertung der jeweiligen Produkte im Beobachtungszeitraum. Sowohl negative als auch Nullwerte wurden aus dieser Spalte entfernt, um nur positive Bewertungsveränderungen in die Analyse miteinzubeziehen, da sonst die Gefahr besteht, dass sich die beobachteten Effekte gegenseitig aufheben. Nullwerte wurden

Tabelle 1: Deskriptive Statistiken ausgewählter Variablen des finalen Datensatzes (gerundet)

ratingdifferenz		n3_add		n3_del	
Min.:	0,10	Min.:	0,00	Min.:	-543,00
1. Quantil:	0,20	1. Quantil:	0,00	1. Quantil:	-0,06
Median:	0,40	Median:	0,00	Median:	0,00
Mean:	0,50	Mean:	0,57	Mean:	-1,07
3. Quantil:	0,60	3. Quantil:	0,09	3. Quantil:	0,00
Max.:	4,00	Max.:	348,00	Max.:	0,00

t3_add		t3_del	
Min.:	0,00	Min.:	0,00
1. Quantil:	0,00	1. Quantil:	0,00
Median:	0,00	Median:	0,00
Mean:	0,19	Mean:	0,13
3. Quantil:	0,39	3. Quantil:	0,16
Max.:	1,00	Max.:	1,00

entfernt, weil nur Änderungen in dieser Variablen relevante Informationen enthalten. Die Variablen „n3_add“ und „n3_del“ bezeichnen den relativen Einfluss der Bewertungshöhe im Zeitintervall hinzukommender (add) und entfernter (del) Rezensionen. „t3_add“ und „t3_del“ bezeichnen analog das relative Durchschnittsalter dieser Rezensionen.

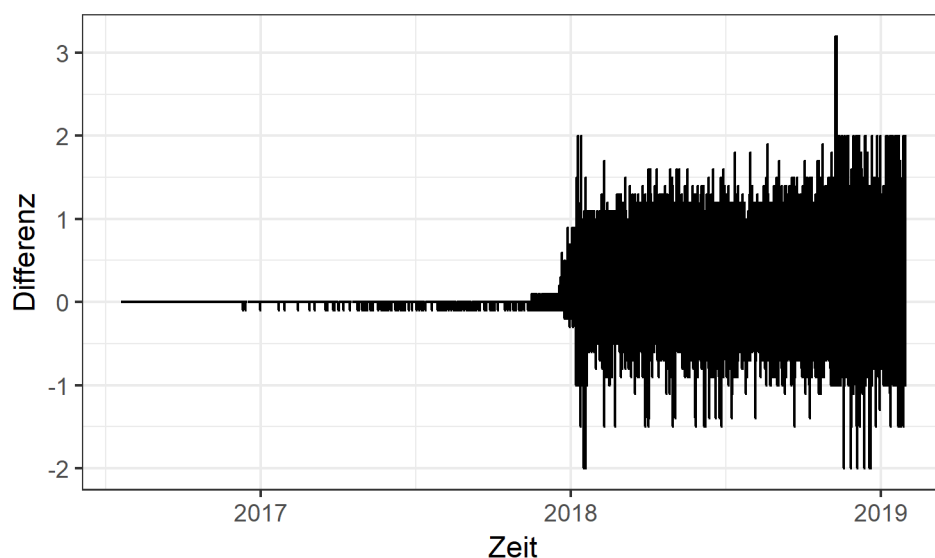
8 Durchführung und Ergebnisse

Nachdem die anzuwendenden Methoden in den vorherigen Kapiteln erläutert wurden, werden diese nun angewendet und ihre Ergebnisse dargestellt. Zunächst wird geprüft, ob es sich bei den Produktgesamtbewertungen um das arithmetische Mittel aller abgegebenen Einzelbewertungen handelt. Anschließend wird der Algorithmus mithilfe der Konzepte des maschinellen Lernens auf eine zeitliche Komponente untersucht.

8.1 Prüfung des Algorithmus auf das arithmetische Mittel

Es gilt zu prüfen, ob es sich bei der Produktgesamtbewertung von Amazon um das arithmetische Mittel aller von Nutzern abgegebenen Einzelbewertungen handelt. Dazu wurde aus den Daten das arithmetische Mittel errechnet. Anschließend wird die Differenz zwischen dem arithmetischen Mittel und tatsächlichem Gesamtwert ermittelt. Es lässt sich erkennen, dass viele Abweichungen des arithmetischen Mittels von Amazons Gesamtbewertung vorliegen. Daher wird in Abbildung 7 neben der auf der Ordinatennachse abgetragenen Differenz

Abbildung 7: Abweichung der Gesamtbewertung vom arithmetischen Mittel der Einzelbewertungen auf Amazon



auf der Abszisse die Zeit ergänzt, sodass die zeitliche Entwicklung ersichtlich ist. Um den Rechenaufwand bei der Erstellung der Grafik in angemessenen Maßen zu halten, wurde eine zufällige Unterstichprobe ($n = 500.000$) aus den vorhandenen Daten gezogen. Es ist eine deutliche zeitliche Entwicklung zu sehen, da vor dem Jahr 2018 nur einige Abweichungen geringen Ausmaßes des arithmetischen Mittels von der tatsächlich von Amazon angegebenen Gesamtbewertung zu sehen sind. Ab Anfang des Jahres 2018 sind hingegen viele starke Differenzen zu beobachten. Es sind mehrere starke Abweichungen von zwei Bewertungspunkten zu sehen. Zu einem Zeitpunkt übersteigt in der Stichprobe die Differenz vom arithmetischen Mittel zur tatsächlichen Gesamtbewertung eine Abweichung von drei Bewertungspunkten. Die starken Abweichungen weisen zunächst augenscheinlich kein Muster auf. Nachfolgend sollen diese Abweichungen mithilfe eines zeitlichen Faktors erklärt werden.

8.2 Trainieren der verschiedenen Modelle des maschinellen Lernens

Aufgrund der im vorherigen Abschnitt gewonnenen Erkenntnisse werden nun nur Daten berücksichtigt, die ab 2018 aufgezeichnet wurden. Daraus ergibt sich die Datensatzgröße von $n = 6.238$. Bevor die finale Analyse beginnen kann, muss der optimale Ansatz des maschinellen Lernens zur Untersuchung des vorliegenden Problems identifiziert werden. Dazu werden Probeläufe mit allen vier vorgestellten Ansätzen gestartet. Jeder Ansatz wird dazu mehrfach mit verschiedenen Einstellungsparametern durchgeführt. Dieser Vorgang kann nur mit begrenzter Variation der Parameter vollzogen werden. Das sogenannte Prüfungsgitter wird dabei so eng wie möglich gefasst. Jedoch muss zugunsten der Rechenzeit eine gewisse Grobheit des Gitters zugelassen werden. Aus diesem Grund werden für diejenigen Ansätze mit der höchsten Vorhersagekraft anschließend erneute Untersuchungen mit einem feineren Gitter im relevanten Intervall der Einstellungsparameter durchgeführt. Zur Validierung wird eine zehnfache Kreuzvalidierung eingesetzt. Die Ergebnisse der Probeläufe werden anhand ihrer Vorhersagegenauigkeit evaluiert und die optimalen Einstellungsparameter festgelegt. Die genauesten Modelle jedes Ansatzes werden anschließend miteinander verglichen, um die Entscheidung für das endgültige Modell zu treffen. Tabelle 2 zeigt die ermittelten optimalen Einstellungsparameter für die unterschiedlichen geprüften Modelle. An dieser Stelle werden ausschließlich die Parameter der besten Modelle jedes Ansatzes dargestellt. Die Einstellungen und Ergebnisse aller Probeläufe sind im Anhang (Tabellen 5-10) zu finden. Aufgrund

Tabelle 2: Optimale Einstellungsparameter der trainierten Modelle

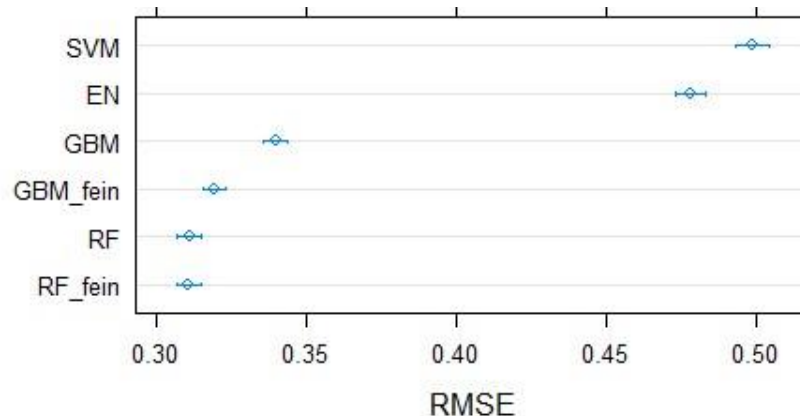
SVM		GBM		fein
Grad:	2	Anzahl Bäume:	150	225
Scale:	0,001	Interaktionstiefe:	3	9
C:	0,5	Lernrate:	0,1	0,1

EN		RF		fein
fraction:	0,525	mtry:	11	9
lambda:	0,1			

der hohen Vorhersagekraft des GBM- und des Zufallswald-Ansatzes werden diese Ansätze erneut mit einem engeren Prüfungsgitter (fein) der Einstellungsparameter geprüft. Die besten Konfigurationen dieser Probeläufe sind ebenfalls in Tabelle 2 aufgeführt.

Das beste Modell wird anhand des RMSE ausgewählt. Abbildung 8 zeigt grafisch den RMSE des präzisesten Modells jedes Ansatzes. Der Zusatz „fein“ in der Modellbezeichnung signalisiert, dass dieses Modell erneut mit einem feineren Prüfungsgitter entwickelt wurde. Die Abbildung verdeutlicht die Rangfolge der Güte der einzelnen Ansätze des maschinellen Lernens. Die exakten RMSE-Werte jedes Modells sind dem Anhang dieser Arbeit (Tabelle 11) zu entnehmen. Es ist zu erkennen, dass der Ansatz der Zufallswälder das beste Ergebnis liefert. Dementsprechend wird dieser Ansatz mit den zugehörigen Einstellungsparametern als finales Modell gewählt und weiter untersucht. Dazu wird das Zufallswald-Modell mit den Einstellungsparametern gebildet.

Abbildung 8: RMSE mit 95%-Konfidenzintervall der Modelle mit den besten Einstellungsparametern



8.3 Das Zufallswald-Modell

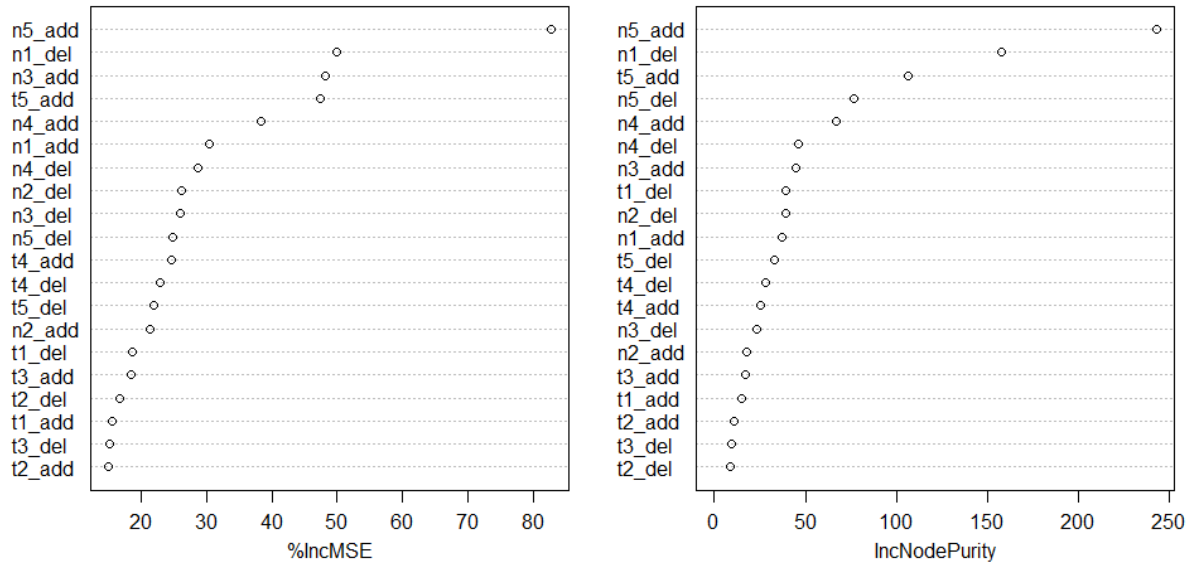
Im vorherigen Abschnitt konnte der Ansatz des Zufallswaldes des maschinellen Lernens als präzisestes Modell identifiziert werden. Von diesem Ansatz ausgehend wird nun das endgültige Modell zur Untersuchung des Gesamtbewertungsalgorithmus von Amazon hinsichtlich einer Zeitgewichtung entwickelt. Dazu werden 500 Entscheidungsbäume mit jeweils neun möglichen abhängigen Variablen pro Knoten gepflanzt. Eine Zusammenfassung des Modellergebnisses zeigt Tabelle 3. Die Ergebniswerte sind im Sinne der Übersicht gerundet worden. Zur Nachvollziehbarkeit des Modells im zur Verfügung gestellten R-Skript wurden die verwendeten Datensätze angegeben. Durch den Zufallswald kann 56,37% der Varianz in den Testdaten erklärt werden. Dieser Wert ist etwas geringer als die erklärte Varianz im Trainingsdatensatz.

Tabelle 3: Modellgesamtergebnis des Zufallswaldes

Formel:	ratingdifferenz ~ .
Daten:	MLData_train
x Testdaten:	MLData_test[, -1]
y Testdaten:	MLData_test[, 1]
Betrachtete Variablen pro Split:	9
Anzahl der Bäume:	500
Durchschnitt der quadrierten Residuen:	0,0974
erklärte Varianz:	60,33%
MSE der Testdaten:	0.1
erklärte Varianz in den Testdaten:	56,37%

Aus diesem Modell können die relativen Wichtigkeiten der einzelnen Variablen abgeleitet werden. Diese sind in Abbildung 9 dargestellt. Hierbei steht es im Vordergrund, die Wichtigkeiten der einzelnen Variablen in Relation zueinander zu betrachten und zu interpretieren. Aus diesem Grund wurde hier eine grafische Aufbereitung der Werte gewählt. Die exakten Werte können dem Anhang dieser Arbeit (Tabelle 12) entnommen werden. Die Darstellung auf der linken Seite zeigt die Wichtigkeit der unabhängigen Variablen anhand des prozentualen Anstiegs des mittleren quadratischen Fehlers, der durch Streichen der Variable aus dem Modell entstünde. Die rechte Grafik zeigt die Wichtigkeit der unabhängigen Variablen gemessen am Zuwachs der Knotenreinheit durch die entsprechenden Variablen. Die Knotenreinheit gibt an, in welchem Maße die Knoten eine eindeutige Entscheidung treffen können (Burger, 2018, S. 136f.). Eine Variable, die eine Veränderung der Anzahl der Rezensionen

Abbildung 9: Wichtigkeiten der unabhängigen Variablen im Zufallswald-Modell



in der jeweiligen Bewertungshöhe angibt, ist mit einem vorangestellten „n“ markiert, gefolgt von der Bewertungshöhe. Eine Variable, die hingegen das Durchschnittsalter angibt, ist analog mit einem vorangestellten „t“ gekennzeichnet. Zusätzlich gibt der Name der Variable Aufschluss darüber, ob sie Zugänge („add“) oder Abgänge („del“) misst. Es ist zu beobachten, dass die Anzahl der hinzukommenden Bewertungen jeder Bewertungshöhe einen größeren Einfluss haben als das durchschnittliche Alter der jeweiligen Rezensionen. Dabei ist anhand des Maßes der Verbesserung der Knotenreinheit des Modells zu sehen, dass das Hinzufügen von extremen, positiven Bewertungen und das Entfernen von extremen, negativen Bewertungen die Spitze der Rangfolge der Wichtigkeit bilden. Insbesondere der Zuwachs von Rezensionen mit der Bewertungshöhe fünf wird durch beide Wichtigkeitsmaße als am einflussreichsten identifiziert. Würde diese Variable im Modell nicht berücksichtigt werden, stiege der MSE um 82,73%. Das Durchschnittsalter dieser Rezensionen weist eine hohe Steigerung der Knotenreinheit auf. Ebenso würde das Weglassen der Variable für die Anzahl der gelöschten Rezensionen der Bewertungshöhe eins für einen Anstieg des MSE um 49,97% sorgen. Am unteren Ende der Wichtigkeitsrangfolge finden sich vor allem die Variablen für das Durchschnittsalter von moderaten (drei) und negativen (eins und zwei) Bewertungen wieder. Diese Variablen haben jeweils eine Wichtigkeit von unter 20%IncMSE.

9 Diskussion

Die Ergebnisse dieser Arbeit zeigen eine Veränderung des Gesamtbewertungsalgorithmus von Amazon. Während zunächst das arithmetische Mittel den Kunden und Informationssuchenden als Gesamtbewertung zur Verfügung gestellt wurde, ist seit Januar 2018 eine Abweichung von diesem arithmetischen Mittel zu beobachten. Wie zu Beginn der Arbeit erörtert, gibt es einige Argumente dafür, einen Gesamtbewertungsalgorithmus mit Elementen, die über das arithmetische Mittel hinausgehen, zu erweitern. Die Gesamtbewertung hat zum Ziel, potenzielle Kunden in ihrem Informationssuchprozess zu unterstützen und Vertrauen zwischen den Marktteilnehmern zu schaffen. Dementsprechend muss dieser Wert zum einen interpretierbar, aber auch aussagekräftig sein. Dieser Arbeit vorausgehende Literatur zeigte, dass das arithmetische Mittel Mängel in seiner Aussagekraft aufweist. Dieses Maß ist anfällig gegenüber Manipulationen und birgt daher die Gefahr, dass die Gesamtbewertung eine deutliche Abweichung von der tatsächlichen Wahrnehmung der Kunden aufweist. Käufer haben dabei das Risiko, ein Produkt zu erhalten, das ihre individuellen Anforderungen nicht bestmöglich erfüllt, da sie nicht repräsentative Bewertungen in ihre Produktevaluation miteinbezogen haben. Darüber hinaus laufen Anbieter Gefahr, von der Konkurrenz manipuliert oder von nicht wohlgesonnenen Geschäftspartnern übermäßig diskreditiert zu werden. Zusätzlich schwächt das arithmetische Mittel als Gesamtbewertung Anreize, die angebotene Qualität der verkauften Produkte zu verbessern.

Nur durch eine Änderung des Algorithmus der Gesamtbewertung lässt sich eine plötzliche Abweichung des Wertes vom arithmetischen Mittel erklären. Amazon scheint auf die Anfälligkeit ihres Bewertungsalgorithmus reagiert und einen neuen Algorithmus entwickelt zu haben, der im Januar 2018 auf der Plattform implementiert wurde. Nachdem ein neuer Algorithmus identifiziert werden konnte, hat es sich diese Arbeit zum Ziel gemacht, neue Elemente des Algorithmus zu ermitteln. Über diese Elemente konnten aus der Literatur einige Hypothesen entwickelt werden. Dazu gehörten Überlegungen zu einer zeitlichen Gewichtung der jeweiligen Rezensionen. Demnach würden Rezensionen, die vor einigen Monaten geschrieben wurden, weniger Auswirkungen auf die Gesamtbewertung haben als aktuelle Rezensionen. Mithilfe des Einsatzes von maschinellem Lernen konnten wesentliche einflussreiche Elemente des Algorithmus identifiziert werden. Ein Vergleich der Ergebnisse

führte zu dem Schluss, dass ein Zufallswald die besten Ergebnisse unter den betrachteten Ansätzen liefern kann.

Es wurde untersucht, wie neue Rezensionen auf die Veränderung der Gesamtbewertung wirken. Dabei konnte gezeigt werden, dass die Höhe der neuen Rezensionen eine große Wirkung auf den Gesamtwert haben. Zudem ist es auffällig, dass hohe Bewertungen einen größeren relativen Einfluss haben als niedrigere. Dieses Phänomen lässt sich durch das Design der Studie erklären. In der Studie wurde nicht der absolute Wert der Gesamtbewertungen als abhängige Variable verwendet, sondern die maximale Veränderung im Beobachtungszeitraum. Durch die größere Differenz zur Gesamtbewertung wirken extreme Bewertungen stärker auf den Gesamtwert als moderate Bewertungen. Eine weitere Eigenschaft der Untersuchung ist die ausschließliche Betrachtung von positiven Entwicklungen über die Zeit. Daher wurden nur Daten über Produkte berücksichtigt, die einen möglichst positiven Aufwärtstrend über die Zeit aufzeigen. Es liegt nahe, dass im relevanten Zeitraum tendenziell mehr positive Bewertungen abgegeben wurden als negative. Deswegen konnte vor allem der Einfluss der Anzahl der hinzukommenden Rezensionen der Bewertungshöhe fünf als besonders stark eingestuft werden. Der umgekehrte Einfluss ist dabei bei den gelöschten Rezensionen zu beobachten. Wird eine Rezension der Bewertungshöhe eins gelöscht, hat dies einen starken Einfluss auf positive Gesamtbewertungsveränderungen. Diese Beobachtung lässt sich genauso über die Extremität der Bewertung und das Studiendesign erklären.

Weiterhin konnte gezeigt werden, dass eine zeitliche Gewichtung der Rezensionen stattfindet. Das relative Durchschnittsalter der Rezensionen der Bewertungshöhe fünf konnte mithilfe des Zufallswaldes als einflussreiches Element im Algorithmus identifiziert werden (47,42%IncMSE). Auch das Durchschnittsalter der Rezensionen der Bewertungshöhe vier sorgt für eine deutliche Senkung des MSE. Durch die Wahl des Studiendesigns lassen sich an dieser Stelle keine Rückschlüsse auf den zeitlichen Einfluss negativer und moderater Bewertungen ziehen. Jedoch weisen die Ergebnisse eine deutliche Tendenz zu einer zeitlichen Gewichtung der Rezensionen im neuen Gesamtbewertungsalgorithmus von Amazon auf. Es ist nicht davon auszugehen, dass im Algorithmus ausschließlich die hohen Bewertungen eine zeitliche Gewichtung erhalten. Die Vermutung liegt nahe, dass sich mit dem umgekehrten Studiendesign für negative Rezensionen eine solche zeitliche Gewichtung ebenfalls nachweisen lässt. Umgekehrtes Studiendesign meint an dieser Stelle die ausschließliche

Betrachtung von Zeiträumen, in denen die Gesamtbewertung der Produkte auf Amazon gefallen ist. Auf die kombinierte Betrachtung der beiden Designs wurde in dieser Arbeit zunächst verzichtet, um eventuell stattfindendes gegenseitiges Aufheben der Effekte zu vermeiden.

Auf Grundlage der Ergebnisse dieser Arbeit lässt sich feststellen, dass Amazon seit Anfang des Jahres 2018 einen neuen Algorithmus zur Berechnung der Produktgesamtbewertungen implementiert hat. Dieser Algorithmus weicht vom arithmetischen Mittel aller abgegebenen Rezensionen ab und wurde durch eine zeitliche Komponente ergänzt. Da mithilfe des Alters der Rezension nur etwa 56% der Varianz in der Bewertungsveränderung über die Zeit erklärt werden kann, ist davon auszugehen, dass der Algorithmus weitere Elemente beinhaltet. Diese Hypothese wird unterstützt von Informationen, die auf der Amazon-Internetseite zu finden sind. Während des Schreibens dieser Bachelorarbeit konnte folgender Text der Seite entnommen werden, über den es vor Anmeldung der Arbeit noch keine Kenntnis gab:

„Amazon berechnet die Sternbewertungen eines Produkts mithilfe eines maschinell gelernten Modells anstelle des Durchschnitts der Rohdaten. Das maschinell gelernte Modell berücksichtigt Faktoren wie das Alter einer Bewertung, die Beurteilung der Nützlichkeit durch Kunden und ob die Bewertungen aus geprüften Einkäufen stammen.“ (Amazon.de [4])

Diese Veröffentlichung unterstützt die Hypothesen und Ergebnisse dieser Arbeit. Neben einer zeitlichen Gewichtung konnte auf Grundlage vorangegangener Literatur eine Gewichtung der Rezensionen nach ihrer Nützlichkeit vermutet werden. Die zeitliche Gewichtung konnte in dieser Arbeit empirisch nachgewiesen werden.

10 Limitationen

Die in dieser Arbeit verwendeten Daten entstammen einer automatisierten Aufzeichnung. Diese Aufzeichnung ist genügend mit einem exakten Zeitstempel und der entsprechenden Zuordnung zum Produkt in Form seiner ASIN dokumentiert. Durch die Automatisierung und die lückenlose Dokumentation kann davon ausgegangen werden, dass die Daten zuverlässig und repräsentativ sind. Die Struktur und Ausprägungen der Daten stellen sich als plausibel heraus. Die Wertebereiche und Werteverteilungen sind unauffällig. Jedoch konnten zu Beginn dieser Bachelorarbeit bei vielen Datenpunkten fehlende Werte ausgemacht werden.

Für diese Datenpunkte musste die Verwendung der Daten in den Analysen korrigiert werden.

Für die Durchführung der Analysen wurde mit R auf einem Computer mit einem Intel® Core™ i7-6650U (2,2GHz, 64 Bit) und 16GB RAM gearbeitet. Die Bearbeitung der Daten und die Durchführung der Analysen nahmen über die Programmierzeit hinaus mehrere Tage Rechenzeit in Anspruch. Aus diesem Grund und unter Berücksichtigung der begrenzten vorgegebenen Bearbeitungszeit dieser Arbeit konnten keine weiterführenden Analysen durchgeführt werden. Insbesondere musste daher auf eine kombinierte Analyse von zeitlich fallender und steigender Entwicklung der Gesamtbewertung verzichtet werden.

Die Analysen weisen zwar eine zeitliche Gewichtung der Rezensionen im Gesamtbewertungsalgorithmus von Amazon auf. Allerdings können dabei keine Rückschlüsse auf die Verknüpfung im Algorithmus oder die exakte Diskontierung gezogen werden.

11 Weiterführende Forschungsansätze

Im Forschungsdesign der Analysen dieser Arbeit wurde der Fokus auf im Beobachtungszeitraum möglichst stark steigende Gesamtbewertungen gelegt. Dadurch konnten auf Grundlage der Ergebnisse nur Rückschlüsse auf die Gewichtung von Rezensionen der Bewertungshöhe vier und fünf gezogen werden. In weiterführenden Analysen gilt es zu prüfen, ob das gleiche Phänomen auch bei moderaten und geringen Bewertungen der Fall ist. Dazu erscheint es sinnvoll, das hier verwendete Forschungsdesign umzukehren und fallende Veränderungen der Produktgesamtbewertung zu betrachten.

Die Ergebnisse zeigen zusätzlich, dass es mindestens einen weiteren Einflussfaktor im Algorithmus von Amazon geben muss. Über die zeitliche Komponente der Rezensionen im Zusammenspiel mit der Bewertungshöhe konnten lediglich etwa 56% der Varianz erklärt werden. Das weist auf weitere wichtige Einflussfaktoren hin, die es in weiteren Forschungen zu identifizieren gilt. Es liegt die Vermutung nahe, dass die von Kunden wahrgenommene Nützlichkeit der Rezensionen eine Rolle spielt. Diese Hypothese wurde durch Informationen auf der Amazon-Internetseite bestärkt. Darüber hinaus muss geprüft werden, ob Informationen über Rezensenten einen gewichtigen Einfluss haben. Dazu können Informationen über die Verifikation der Käufer und Bewertungen der Rezensenten herangezogen werden.

Sind alle Einflussfaktoren des Bewertungsalgorithmus identifiziert worden, können die Informationen genutzt werden, um den Algorithmus auf die Verknüpfung der Faktoren zu prüfen. Daraufhin können die exakten Parameter der Berechnung ermittelt werden. Ist dieser Schritt vollzogen, wurde der gesamte Algorithmus rekonstruiert und kann von allen betroffenen Parteien genutzt werden. Kunden erhalten Transparenz über den Wert, den sie zur Bewertungszusammenfassung angezeigt bekommen, und können selbst entscheiden, inwiefern sie diesem Ergebnis Vertrauen schenken. Anbieter haben die Möglichkeit, sich über die gewährte Sicherheit vor Manipulationen des Algorithmus zu informieren. Darüber hinaus können Verkäufer Schwachstellen des Algorithmus identifizieren und für ihren Vorteil im Sinne der Absatzsteigerung nutzen. Andere Online-Plattformen können den Algorithmus in ihrem eigenen Rezensionssystem implementieren, sowie eigene Ressourcen aufwenden, um ihn weiterzuentwickeln.

Zusätzlich gibt es Forschungsbedarf hinsichtlich der Wirkung des neuen Algorithmus. Insbesondere die Wirkung auf den Absatz sollte zentraler Bestandteil zukünftiger Studien sein. Darüber hinaus kann die Entwicklung manipulativer Handlungen unter dem neuen Algorithmus untersucht werden.

12 Fazit

Rezensionssysteme gehören heutzutage zum Grundbaustein einer Online-Handels-Plattform. Viele kleinere Online-Händler kaufen die Rezensionen für ihre Produkte von Amazon, um sie auf der eigenen Seite einzubetten (Mudambi/Schuff, 2010). So angesehen Amazons Rezensionssystem ist, ergibt sich ein vielfältiger wirtschaftlicher Anreiz, es fortlaufend weiterzuentwickeln. Mit der Wirkung und Größe des Systems steigt auch seine strategische Bedeutung und damit einhergehend die Gefahr vor Manipulationen. Amazon scheint diese Gefahr erkannt zu haben und implementierte ein Update für das Rezensionssystem, durch welches sich die Berechnung der Produktgesamtbewertungen deutlich änderte. Diese Arbeit zeigt, dass diese Gesamtbewertungen bis Ende 2017 dem arithmetischen Mittel aller für das Produkt abgegebenen Bewertungen entsprachen. Ab Anfang 2018 konnte eine deutliche Abweichung von diesem Berechnungsalgorithmus identifiziert werden.

Auf Grundlage eines umfangreichen Datensatzes aus dem Rezensionssystem von Amazon konnte der neue Bewertungsalgorithmus auf seine enthaltenen Elemente untersucht werden.

Dabei lag der Fokus der Untersuchungen auf dem Nachweis einer zeitlichen Komponente im Algorithmus, der die Rezensionen hinsichtlich ihres Alters in der Gesamtbewertung unterschiedlich gewichtet. Dazu wurde der Ansatz des maschinellen Lernens gewählt. Zur Bestimmung des besten Modelltyps wurden einige Probeläufe durchgeführt. Es wurden Elastische Netze, Stützvektormaschinen, Gradient Boosting Maschinen und Zufallswälder mit verschiedenen Einstellungsparametern trainiert und hinsichtlich ihrer Güte beurteilt. Ein Zufallswald-Modell erwies sich dabei als das präziseste. Mithilfe des Modells konnte gezeigt werden, dass das Alter der Rezensionen ein entscheidendes Element des neuen Amazon-Algorithmus darstellt. Zusammen mit der Höhe der abgegebenen Bewertungen konnten mit diesem Element etwa 56% der Varianz im zeitlichen Verlauf der Gesamtbewertungen erklärt werden. Die Ergebnisse weisen zudem darauf hin, dass der neue Amazon-Algorithmus noch weitere Komponenten neben einer zeitlichen Gewichtung und der Bewertungshöhe enthält. Die Komponenten sollen Gegenstand zukünftiger Untersuchungen sein.

Diese Arbeit zeigt einen Ansatz zur Rekonstruktion von Bewertungsalgorithmen. Wissenschaftliche Arbeiten zu diesem Themenkomplex legen Grundlagen zur weiteren Forschung an solchen Algorithmen. Die Forschung und die Vermittlung der Ergebnisse tragen dazu bei, dass Kunden entscheiden können, wie viel Vertrauen sie in solche Rezensionssysteme und ihre aufbereiteten Zusammenfassungen setzen. Zudem können sich Anbieter über die Fairness ihrer Bewertung bewusst werden. Allerdings zeigt diese Forschung auch Ansatzpunkte für zukünftige strategische Manipulationen auf. Daher ist es sinnvoll, eine fortlaufende Entwicklung von Rezensionssystemen anzustreben.

Anhang

Tabelle 4: Datenstruktur der verwendeten Variablen (gerundet)

ratingdifferenz	
Min.:	0,10
1. Quantil:	0,20
Median:	0,40
Mean:	0,50
3. Quantil:	0,60
Max.:	4,00

t1_add	
Min.:	0,00
1. Quantil:	0,00
Median:	0,00
Mean:	0,16
3. Quantil:	0,33
Max.:	1,00

t1_del	
Min.:	0,00
1. Quantil:	0,00
Median:	0,00
Mean:	0,15
3. Quantil:	0,23
Max.:	1,00

n1_add	
Min.:	0,00
1. Quantil:	0,00
Median:	0,00
Mean:	0,18
3. Quantil:	0,02
Max.:	95,00

n1_del	
Min.:	-205,00
1. Quantil:	-0,03
Median:	0,00
Mean:	-0,41
3. Quantil:	0,00
Max.:	0,00

t2_add	
Min.:	0,00
1. Quantil:	0,00
Median:	0,00
Mean:	0,14
3. Quantil:	0,19
Max.:	0,99

t2_del	
Min.:	0,00
1. Quantil:	0,00
Median:	0,00
Mean:	0,11
3. Quantil:	0,00
Max.:	1,00

n2_add	
Min.:	0,00
1. Quantil:	0,00
Median:	0,00
Mean:	0,21
3. Quantil:	0,02
Max.:	114,00

n2_del	
Min.:	-252,00
1. Quantil:	-0,02
Median:	0,00
Mean:	-0,45
3. Quantil:	0,00
Max.:	0,00

t3_add	
Min.:	0,00
1. Quantil:	0,00
Median:	0,00
Mean:	0,19
3. Quantil:	0,39
Max.:	1,00

t3_del	
Min.:	0,00
1. Quantil:	0,00
Median:	0,00
Mean:	0,13
3. Quantil:	0,16
Max.:	1,00

n3_add	
Min.:	0,00
1. Quantil:	0,00
Median:	0,00
Mean:	0,57
3. Quantil:	0,09
Max.:	348,00

n3_del	
Min.:	-543,00
1. Quantil:	-0,06
Median:	0,00
Mean:	-1,07
3. Quantil:	0,00
Max.:	0,00

t4_add	
Min.:	0,00
1. Quantil:	0,00
Median:	0,00
Mean:	0,24
3. Quantil:	0,47
Max.:	1,00

t4_del	
Min.:	0,00
1. Quantil:	0,00
Median:	0,00
Mean:	0,13
3. Quantil:	0,16
Max.:	1,00

n4_add	
Min.:	0,00
1. Quantil:	0,00
Median:	0,05
Mean:	1,60
3. Quantil:	0,37
Max.:	1052,00

n4_del	
Min.:	-1548,00
1. Quantil:	-0,25
Median:	0,00
Mean:	-3,15
3. Quantil:	0,00
Max.:	0,00

t5_add	
Min.:	0,00
1. Quantil:	0,11
Median:	0,37
Mean:	0,33
3. Quantil:	0,49
Max.:	1,00

t5_del	
Min.:	0,00
1. Quantil:	0,00
Median:	0,00
Mean:	0,22
3. Quantil:	0,45
Max.:	0,99

n5_add	
Min.:	0,00
1. Quantil:	0,62
Median:	1,37
Mean:	9,66
3. Quantil:	2,50
Max.:	6670,00

n5_del	
Min.:	-9195,00
1. Quantil:	-0,42
Median:	-0,05
Mean:	-20,31
3. Quantil:	0,00
Max.:	0,00

Tabelle 5: Einstellungen und Ergebnisse der GBM-Probelaufe (gerundet)

Lernrate	Interaktionstiefe	n.trees	RMSE	Rsquared	MAE
0,1	1	50	0,4227	0,2809	0,2718
0,1	1	100	0,4113	0,3122	0,2662
0,1	1	150	0,4053	0,3286	0,2637
0,1	2	50	0,3830	0,4174	0,2515
0,1	2	100	0,3631	0,4650	0,2392
0,1	2	150	0,3532	0,4888	0,2326
0,1	3	50	0,3653	0,4667	0,2424
0,1	3	100	0,3475	0,5057	0,2300
0,1	3	150	0,3398	0,5239	0,2234

Stochastic Gradient Boosting

6238 Stichproben

20 Prädiktoren

Keine Vorverarbeitung

Resampling: Kreuzvalidierung (10-fach, 10 Wiederholungen)

Stichprobengrößen: 5615, 5615, 5614, 5614, 5613, 5613, ...

Tabelle 6: Einstellungen und Ergebnisse der RF-Probelaufe (gerundet)

mtry	RMSE	Rsquared	MAE
2	0,3303	0,5687	0,1955
11	0,3111	0,5993	0,1725
20	0,3139	0,5919	0,1728

Random Forest

6238 Stichproben

20 Prädiktoren

Keine Vorverarbeitung

Resampling: Kreuzvalidierung (10-fach, 10 Wiederholungen)

Stichprobengrößen: 5613, 5613, 5615, 5615, 5613, 5615, ...

Tabelle 7: Einstellungen und Ergebnisse der RF-Probelaufe mit feinem Gitter (gerundet)

mtry	RMSE	Rsquared	MAE
8	0,3110	0,6001	0,1736
9	0,3108	0,6003	0,1731
10	0,3110	0,5996	0,1727
11	0,3113	0,5988	0,1725

Random Forest

6238 Stichproben

20 Prädiktoren

Keine Vorverarbeitung

Resampling: Kreuzvalidierung (10-fach, 10 Wiederholungen)

Stichprobengrößen: 5613, 5613, 5615, 5615, 5613, 5615, ...

Tabelle 8: Einstellungen und Ergebnisse der EN-Probelaufe (gerundet)

lambda	fraction	RMSE	Rsquared	MAE
0E+00	0,05	0,4836	0,0435	0,3248
0E+00	0,525	0,4815	0,0626	0,3234
0E+00	1	0,4847	0,0598	0,3238
1E-04	0,05	0,4836	0,0432	0,3248
1E-04	0,525	0,4814	0,0626	0,3234
1E-04	1	0,4846	0,0600	0,3237
1E-01	0,05	0,4884	0,0356	0,3274
1E-01	0,525	0,4781	0,0620	0,3219
1E-01	1	0,4794	0,0641	0,3232

Elasticnet

6238 Stichproben

20 Prädiktoren

Keine Vorverarbeitung

Resampling: Kreuzvalidierung (10-fach, 10 Wiederholungen)

Stichprobengrößen: 5613, 5613, 5615, 5615, 5613, 5615, ...

Tabelle 9: Einstellungen und Ergebnisse der SVM-Probelaufe (gerundet)

Grad	scale	C	RMSE	Rsquared	MAE
1	0,001	0,25	0,5039	0,0531	0,2994
1	0,001	0,5	0,5014	0,0565	0,2986
1	0,001	1	0,5010	0,0584	0,2984
1	0,01	0,25	0,5035	0,0599	0,2984
1	0,01	0,5	0,5055	0,0597	0,2982
1	0,01	1	0,5071	0,0593	0,2978
1	0,1	0,25	0,5090	0,0586	0,2974
1	0,1	0,5	0,5110	0,0580	0,2974
1	0,1	1	0,5126	0,0580	0,2973
2	0,001	0,25	0,4993	0,0599	0,2977
2	0,001	0,5	0,4992	0,0604	0,2974
2	0,001	1	0,5051	0,0629	0,2970
2	0,01	0,25	0,6648	0,0899	0,2963
2	0,01	0,5	0,7072	0,0928	0,2953
2	0,01	1	0,7522	0,0957	0,2957
2	0,1	0,25	0,8005	0,0890	0,2962
2	0,1	0,5	0,7715	0,0843	0,2952
2	0,1	1	0,7779	0,0795	0,2961
3	0,001	0,25	0,5306	0,0585	0,2993
3	0,001	0,5	0,5740	0,0634	0,3003
3	0,001	1	0,6404	0,0682	0,3016
3	0,01	0,25	1,5707	0,0773	0,3312
3	0,01	0,5	3,5209	0,0752	0,4080
3	0,01	1	6,2369	0,0737	0,5170
3	0,1	0,25	12,5962	0,0547	0,7943
3	0,1	0,5	14,6110	0,0537	0,8970
3	0,1	1	18,3010	0,0462	1,0775

Support Vector Machines with Polynomial Kernel

6238 Stichproben

20 Prädiktoren

Keine Vorverarbeitung

Resampling: Kreuzvalidierung (10-fach, 10 Wiederholungen)

Stichprobengrößen: 5613, 5613, 5615, 5615, 5613, 5615, ...

Tabelle 10: Einstellungen und Ergebnisse der GBM-Probelaufe mit feinem Gitter (gerundet)

Lernrate	Interaktionstiefe	n.minobsinnode	n.trees	RMSE	Rsquared	MAE
0,075	3	7	150	0,3449	0,5120	0,2279
0,075	3	7	175	0,3421	0,5185	0,2254
0,075	3	7	200	0,3398	0,5240	0,2233
0,075	3	7	225	0,3379	0,5286	0,2216
0,075	3	10	150	0,3447	0,5124	0,2278
0,075	3	10	175	0,3418	0,5192	0,2254
0,075	3	10	200	0,3395	0,5247	0,2233
0,075	3	10	225	0,3377	0,5292	0,2215
0,075	3	12	150	0,3447	0,5123	0,2278
0,075	3	12	175	0,3419	0,5190	0,2253
0,075	3	12	200	0,3398	0,5243	0,2232
0,075	3	12	225	0,3379	0,5289	0,2213
0,075	3	15	150	0,3447	0,5124	0,2277
0,075	3	15	175	0,3418	0,5193	0,2252
0,075	3	15	200	0,3395	0,5249	0,2231
0,075	3	15	225	0,3377	0,5295	0,2213
0,075	4	7	150	0,3376	0,5299	0,2223
0,075	4	7	175	0,3354	0,5353	0,2200
0,075	4	7	200	0,3334	0,5404	0,2179
0,075	4	7	225	0,3321	0,5439	0,2164
0,075	4	10	150	0,3374	0,5306	0,2219
0,075	4	10	175	0,3350	0,5365	0,2195
0,075	4	10	200	0,3332	0,5411	0,2176
0,075	4	10	225	0,3316	0,5451	0,2158
0,075	4	12	150	0,3377	0,5297	0,2220
0,075	4	12	175	0,3353	0,5357	0,2196
0,075	4	12	200	0,3335	0,5405	0,2176
0,075	4	12	225	0,3318	0,5448	0,2158
0,075	4	15	150	0,3380	0,5290	0,2222
0,075	4	15	175	0,3355	0,5353	0,2198
0,075	4	15	200	0,3335	0,5403	0,2178
0,075	4	15	225	0,3320	0,5442	0,2161
0,075	5	7	150	0,3334	0,5404	0,2183
0,075	5	7	175	0,3313	0,5458	0,2160
0,075	5	7	200	0,3298	0,5497	0,2142
0,075	5	7	225	0,3283	0,5537	0,2124
0,075	5	10	150	0,3329	0,5418	0,2179
0,075	5	10	175	0,3308	0,5475	0,2155
0,075	5	10	200	0,3290	0,5522	0,2135
0,075	5	10	225	0,3275	0,5561	0,2118
0,075	5	12	150	0,3328	0,5424	0,2177
0,075	5	12	175	0,3306	0,5481	0,2154
0,075	5	12	200	0,3288	0,5527	0,2134
0,075	5	12	225	0,3274	0,5562	0,2117
0,075	5	15	150	0,3327	0,5426	0,2177
0,075	5	15	175	0,3307	0,5477	0,2154
0,075	5	15	200	0,3290	0,5521	0,2135
0,075	5	15	225	0,3276	0,5558	0,2118
0,075	6	7	150	0,3302	0,5489	0,2152
0,075	6	7	175	0,3282	0,5540	0,2130

Lernrate	Interaktionstiefe	n.minobsinnode	n.trees	RMSE	Rsquared	MAE
0,075	6	7	200	0,3267	0,5580	0,2110
0,075	6	7	225	0,3256	0,5607	0,2094
0,075	6	10	150	0,3298	0,5498	0,2148
0,075	6	10	175	0,3280	0,5546	0,2125
0,075	6	10	200	0,3265	0,5584	0,2106
0,075	6	10	225	0,3253	0,5615	0,2089
0,075	6	12	150	0,3292	0,5519	0,2143
0,075	6	12	175	0,3273	0,5569	0,2120
0,075	6	12	200	0,3258	0,5606	0,2102
0,075	6	12	225	0,3245	0,5641	0,2086
0,075	6	15	150	0,3304	0,5484	0,2147
0,075	6	15	175	0,3284	0,5536	0,2125
0,075	6	15	200	0,3270	0,5572	0,2108
0,075	6	15	225	0,3259	0,5602	0,2093
0,075	7	7	150	0,3276	0,5556	0,2125
0,075	7	7	175	0,3257	0,5606	0,2101
0,075	7	7	200	0,3244	0,5640	0,2082
0,075	7	7	225	0,3231	0,5673	0,2064
0,075	7	10	150	0,3275	0,5558	0,2123
0,075	7	10	175	0,3257	0,5604	0,2100
0,075	7	10	200	0,3243	0,5640	0,2082
0,075	7	10	225	0,3233	0,5667	0,2066
0,075	7	12	150	0,3271	0,5570	0,2118
0,075	7	12	175	0,3254	0,5615	0,2096
0,075	7	12	200	0,3240	0,5651	0,2078
0,075	7	12	225	0,3231	0,5674	0,2064
0,075	7	15	150	0,3271	0,5571	0,2122
0,075	7	15	175	0,3254	0,5615	0,2101
0,075	7	15	200	0,3240	0,5653	0,2082
0,075	7	15	225	0,3228	0,5682	0,2067
0,075	8	7	150	0,3257	0,5607	0,2104
0,075	8	7	175	0,3241	0,5647	0,2081
0,075	8	7	200	0,3229	0,5678	0,2063
0,075	8	7	225	0,3220	0,5703	0,2047
0,075	8	10	150	0,3251	0,5624	0,2097
0,075	8	10	175	0,3236	0,5663	0,2076
0,075	8	10	200	0,3224	0,5694	0,2057
0,075	8	10	225	0,3214	0,5719	0,2042
0,075	8	12	150	0,3253	0,5619	0,2099
0,075	8	12	175	0,3238	0,5657	0,2078
0,075	8	12	200	0,3226	0,5688	0,2061
0,075	8	12	225	0,3216	0,5715	0,2044
0,075	8	15	150	0,3258	0,5604	0,2101
0,075	8	15	175	0,3241	0,5648	0,2079
0,075	8	15	200	0,3229	0,5680	0,2062
0,075	8	15	225	0,3218	0,5710	0,2046
0,075	9	7	150	0,3248	0,5630	0,2088
0,075	9	7	175	0,3233	0,5668	0,2065
0,075	9	7	200	0,3223	0,5697	0,2047
0,075	9	7	225	0,3213	0,5722	0,2031
0,075	9	10	150	0,3244	0,5640	0,2084
0,075	9	10	175	0,3228	0,5682	0,2061

Lernrate	Interaktionstiefe	n.minobsinnode	n.trees	RMSE	Rsquared	MAE
0,075	9	10	200	0,3216	0,5713	0,2043
0,075	9	10	225	0,3207	0,5738	0,2026
0,075	9	12	150	0,3241	0,5649	0,2083
0,075	9	12	175	0,3228	0,5685	0,2062
0,075	9	12	200	0,3217	0,5712	0,2044
0,075	9	12	225	0,3207	0,5737	0,2027
0,075	9	15	150	0,3244	0,5643	0,2085
0,075	9	15	175	0,3230	0,5680	0,2065
0,075	9	15	200	0,3218	0,5709	0,2046
0,075	9	15	225	0,3209	0,5734	0,2031
0,1	3	7	150	0,3404	0,5219	0,2239
0,1	3	7	175	0,3382	0,5276	0,2217
0,1	3	7	200	0,3364	0,5321	0,2197
0,1	3	7	225	0,3345	0,5370	0,2179
0,1	3	10	150	0,3400	0,5235	0,2232
0,1	3	10	175	0,3376	0,5294	0,2207
0,1	3	10	200	0,3356	0,5345	0,2187
0,1	3	10	225	0,3342	0,5383	0,2172
0,1	3	12	150	0,3399	0,5238	0,2233
0,1	3	12	175	0,3375	0,5297	0,2209
0,1	3	12	200	0,3356	0,5348	0,2189
0,1	3	12	225	0,3341	0,5388	0,2172
0,1	3	15	150	0,3399	0,5236	0,2232
0,1	3	15	175	0,3375	0,5298	0,2209
0,1	3	15	200	0,3356	0,5346	0,2190
0,1	3	15	225	0,3340	0,5387	0,2173
0,1	4	7	150	0,3344	0,5376	0,2187
0,1	4	7	175	0,3325	0,5423	0,2165
0,1	4	7	200	0,3308	0,5468	0,2146
0,1	4	7	225	0,3295	0,5502	0,2129
0,1	4	10	150	0,3343	0,5380	0,2183
0,1	4	10	175	0,3321	0,5438	0,2159
0,1	4	10	200	0,3305	0,5480	0,2140
0,1	4	10	225	0,3288	0,5524	0,2123
0,1	4	12	150	0,3340	0,5388	0,2180
0,1	4	12	175	0,3318	0,5444	0,2157
0,1	4	12	200	0,3301	0,5490	0,2136
0,1	4	12	225	0,3287	0,5528	0,2120
0,1	4	15	150	0,3340	0,5389	0,2182
0,1	4	15	175	0,3320	0,5439	0,2160
0,1	4	15	200	0,3304	0,5483	0,2141
0,1	4	15	225	0,3291	0,5515	0,2125
0,1	5	7	150	0,3304	0,5479	0,2146
0,1	5	7	175	0,3285	0,5528	0,2125
0,1	5	7	200	0,3272	0,5565	0,2106
0,1	5	7	225	0,3259	0,5599	0,2089
0,1	5	10	150	0,3301	0,5491	0,2142
0,1	5	10	175	0,3279	0,5550	0,2119
0,1	5	10	200	0,3265	0,5585	0,2100
0,1	5	10	225	0,3253	0,5616	0,2084
0,1	5	12	150	0,3300	0,5493	0,2141
0,1	5	12	175	0,3281	0,5543	0,2119

Lernrate	Interaktionstiefe	n.minobsinnode	n.trees	RMSE	Rsquared	MAE
0,1	5	12	200	0,3265	0,5585	0,2100
0,1	5	12	225	0,3253	0,5617	0,2085
0,1	5	15	150	0,3294	0,5508	0,2138
0,1	5	15	175	0,3276	0,5555	0,2118
0,1	5	15	200	0,3262	0,5592	0,2101
0,1	5	15	225	0,3249	0,5626	0,2084
0,1	6	7	150	0,3275	0,5558	0,2113
0,1	6	7	175	0,3259	0,5600	0,2091
0,1	6	7	200	0,3248	0,5627	0,2074
0,1	6	7	225	0,3239	0,5654	0,2059
0,1	6	10	150	0,3268	0,5576	0,2107
0,1	6	10	175	0,3252	0,5618	0,2085
0,1	6	10	200	0,3239	0,5651	0,2069
0,1	6	10	225	0,3229	0,5679	0,2053
0,1	6	12	150	0,3271	0,5569	0,2112
0,1	6	12	175	0,3253	0,5617	0,2091
0,1	6	12	200	0,3241	0,5648	0,2074
0,1	6	12	225	0,3230	0,5678	0,2059
0,1	6	15	150	0,3266	0,5583	0,2109
0,1	6	15	175	0,3250	0,5626	0,2089
0,1	6	15	200	0,3238	0,5658	0,2073
0,1	6	15	225	0,3228	0,5683	0,2059
0,1	7	7	150	0,3252	0,5618	0,2087
0,1	7	7	175	0,3238	0,5654	0,2066
0,1	7	7	200	0,3227	0,5686	0,2049
0,1	7	7	225	0,3221	0,5701	0,2035
0,1	7	10	150	0,3247	0,5629	0,2084
0,1	7	10	175	0,3232	0,5669	0,2065
0,1	7	10	200	0,3222	0,5695	0,2047
0,1	7	10	225	0,3212	0,5724	0,2031
0,1	7	12	150	0,3251	0,5620	0,2085
0,1	7	12	175	0,3238	0,5656	0,2065
0,1	7	12	200	0,3228	0,5683	0,2048
0,1	7	12	225	0,3217	0,5712	0,2032
0,1	7	15	150	0,3254	0,5614	0,2091
0,1	7	15	175	0,3239	0,5653	0,2072
0,1	7	15	200	0,3227	0,5685	0,2056
0,1	7	15	225	0,3219	0,5707	0,2040
0,1	8	7	150	0,3239	0,5653	0,2068
0,1	8	7	175	0,3226	0,5689	0,2047
0,1	8	7	200	0,3216	0,5715	0,2028
0,1	8	7	225	0,3209	0,5734	0,2012
0,1	8	10	150	0,3234	0,5665	0,2066
0,1	8	10	175	0,3222	0,5697	0,2047
0,1	8	10	200	0,3213	0,5723	0,2028
0,1	8	10	225	0,3204	0,5748	0,2012
0,1	8	12	150	0,3235	0,5663	0,2066
0,1	8	12	175	0,3222	0,5697	0,2045
0,1	8	12	200	0,3211	0,5727	0,2026
0,1	8	12	225	0,3204	0,5746	0,2012
0,1	8	15	150	0,3235	0,5665	0,2069
0,1	8	15	175	0,3220	0,5702	0,2049

Lernrate	Interaktionstiefe	n.minobsinnode	n.trees	RMSE	Rsquared	MAE
0,1	8	15	200	0,3211	0,5728	0,2032
0,1	8	15	225	0,3203	0,5749	0,2017
0,1	9	7	150	0,3227	0,5683	0,2050
0,1	9	7	175	0,3217	0,5711	0,2029
0,1	9	7	200	0,3209	0,5735	0,2011
0,1	9	7	225	0,3202	0,5754	0,1995
0,1	9	10	150	0,3226	0,5687	0,2047
0,1	9	10	175	0,3214	0,5719	0,2024
0,1	9	10	200	0,3204	0,5749	0,2006
0,1	9	10	225	0,3196	0,5769	0,1991
0,1	9	12	150	0,3229	0,5680	0,2051
0,1	9	12	175	0,3217	0,5713	0,2031
0,1	9	12	200	0,3209	0,5735	0,2014
0,1	9	12	225	0,3200	0,5761	0,1998
0,1	9	15	150	0,3225	0,5691	0,2053
0,1	9	15	175	0,3213	0,5723	0,2034
0,1	9	15	200	0,3204	0,5748	0,2017
0,1	9	15	225	0,3198	0,5765	0,2003
0,125	3	7	150	0,3378	0,5282	0,2207
0,125	3	7	175	0,3359	0,5332	0,2188
0,125	3	7	200	0,3341	0,5379	0,2167
0,125	3	7	225	0,3328	0,5415	0,2152
0,125	3	10	150	0,3371	0,5306	0,2203
0,125	3	10	175	0,3352	0,5356	0,2182
0,125	3	10	200	0,3334	0,5402	0,2164
0,125	3	10	225	0,3319	0,5442	0,2148
0,125	3	12	150	0,3367	0,5315	0,2203
0,125	3	12	175	0,3346	0,5370	0,2181
0,125	3	12	200	0,3328	0,5418	0,2163
0,125	3	12	225	0,3315	0,5452	0,2147
0,125	3	15	150	0,3374	0,5297	0,2204
0,125	3	15	175	0,3355	0,5347	0,2184
0,125	3	15	200	0,3338	0,5392	0,2166
0,125	3	15	225	0,3323	0,5433	0,2150
0,125	4	7	150	0,3321	0,5431	0,2157
0,125	4	7	175	0,3303	0,5481	0,2135
0,125	4	7	200	0,3290	0,5515	0,2117
0,125	4	7	225	0,3277	0,5550	0,2100
0,125	4	10	150	0,3317	0,5446	0,2150
0,125	4	10	175	0,3300	0,5489	0,2129
0,125	4	10	200	0,3286	0,5527	0,2112
0,125	4	10	225	0,3275	0,5556	0,2097
0,125	4	12	150	0,3317	0,5446	0,2152
0,125	4	12	175	0,3298	0,5497	0,2130
0,125	4	12	200	0,3281	0,5541	0,2111
0,125	4	12	225	0,3270	0,5572	0,2096
0,125	4	15	150	0,3315	0,5453	0,2152
0,125	4	15	175	0,3298	0,5498	0,2132
0,125	4	15	200	0,3283	0,5538	0,2115
0,125	4	15	225	0,3270	0,5570	0,2100
0,125	5	7	150	0,3288	0,5522	0,2121
0,125	5	7	175	0,3273	0,5561	0,2100

Lernrate	Interaktionstiefe	n.minobsinnode	n.trees	RMSE	Rsquared	MAE
0,125	5	7	200	0,3263	0,5588	0,2083
0,125	5	7	225	0,3251	0,5620	0,2066
0,125	5	10	150	0,3277	0,5554	0,2111
0,125	5	10	175	0,3260	0,5599	0,2090
0,125	5	10	200	0,3247	0,5633	0,2072
0,125	5	10	225	0,3236	0,5663	0,2057
0,125	5	12	150	0,3273	0,5564	0,2110
0,125	5	12	175	0,3254	0,5614	0,2089
0,125	5	12	200	0,3243	0,5645	0,2073
0,125	5	12	225	0,3231	0,5675	0,2059
0,125	5	15	150	0,3289	0,5520	0,2123
0,125	5	15	175	0,3270	0,5572	0,2101
0,125	5	15	200	0,3257	0,5607	0,2084
0,125	5	15	225	0,3247	0,5634	0,2069
0,125	6	7	150	0,3267	0,5578	0,2092
0,125	6	7	175	0,3252	0,5617	0,2071
0,125	6	7	200	0,3242	0,5645	0,2054
0,125	6	7	225	0,3233	0,5671	0,2039
0,125	6	10	150	0,3257	0,5607	0,2086
0,125	6	10	175	0,3243	0,5645	0,2068
0,125	6	10	200	0,3232	0,5675	0,2051
0,125	6	10	225	0,3224	0,5697	0,2037
0,125	6	12	150	0,3255	0,5609	0,2088
0,125	6	12	175	0,3244	0,5639	0,2069
0,125	6	12	200	0,3233	0,5669	0,2052
0,125	6	12	225	0,3225	0,5693	0,2036
0,125	6	15	150	0,3255	0,5614	0,2088
0,125	6	15	175	0,3242	0,5648	0,2070
0,125	6	15	200	0,3229	0,5682	0,2053
0,125	6	15	225	0,3221	0,5703	0,2038
0,125	7	7	150	0,3249	0,5627	0,2067
0,125	7	7	175	0,3235	0,5662	0,2044
0,125	7	7	200	0,3227	0,5686	0,2027
0,125	7	7	225	0,3223	0,5700	0,2012
0,125	7	10	150	0,3235	0,5664	0,2062
0,125	7	10	175	0,3225	0,5689	0,2043
0,125	7	10	200	0,3216	0,5715	0,2026
0,125	7	10	225	0,3212	0,5727	0,2014
0,125	7	12	150	0,3241	0,5648	0,2065
0,125	7	12	175	0,3228	0,5682	0,2046
0,125	7	12	200	0,3220	0,5706	0,2029
0,125	7	12	225	0,3214	0,5723	0,2014
0,125	7	15	150	0,3236	0,5663	0,2064
0,125	7	15	175	0,3222	0,5700	0,2044
0,125	7	15	200	0,3214	0,5722	0,2029
0,125	7	15	225	0,3208	0,5740	0,2016
0,125	8	7	150	0,3234	0,5666	0,2048
0,125	8	7	175	0,3221	0,5701	0,2026
0,125	8	7	200	0,3217	0,5716	0,2009
0,125	8	7	225	0,3213	0,5728	0,1995
0,125	8	10	150	0,3227	0,5688	0,2044
0,125	8	10	175	0,3218	0,5714	0,2025

Lernrate	Interaktionstiefe	n.minobsinnode	n.trees	RMSE	Rsquared	MAE
0,125	8	10	200	0,3211	0,5733	0,2009
0,125	8	10	225	0,3208	0,5742	0,1995
0,125	8	12	150	0,3223	0,5699	0,2045
0,125	8	12	175	0,3211	0,5731	0,2024
0,125	8	12	200	0,3203	0,5753	0,2007
0,125	8	12	225	0,3199	0,5766	0,1993
0,125	8	15	150	0,3221	0,5702	0,2045
0,125	8	15	175	0,3212	0,5727	0,2025
0,125	8	15	200	0,3204	0,5749	0,2011
0,125	8	15	225	0,3199	0,5764	0,1997
0,125	9	7	150	0,3225	0,5694	0,2031
0,125	9	7	175	0,3218	0,5714	0,2012
0,125	9	7	200	0,3211	0,5735	0,1995
0,125	9	7	225	0,3206	0,5749	0,1981
0,125	9	10	150	0,3217	0,5715	0,2023
0,125	9	10	175	0,3207	0,5743	0,2002
0,125	9	10	200	0,3200	0,5762	0,1986
0,125	9	10	225	0,3196	0,5775	0,1972
0,125	9	12	150	0,3215	0,5719	0,2028
0,125	9	12	175	0,3207	0,5743	0,2008
0,125	9	12	200	0,3201	0,5759	0,1994
0,125	9	12	225	0,3197	0,5772	0,1979
0,125	9	15	150	0,3216	0,5714	0,2031
0,125	9	15	175	0,3207	0,5741	0,2012
0,125	9	15	200	0,3202	0,5756	0,1996
0,125	9	15	225	0,3197	0,5769	0,1982
0,15	3	7	150	0,3361	0,5325	0,2190
0,15	3	7	175	0,3343	0,5372	0,2170
0,15	3	7	200	0,3326	0,5418	0,2153
0,15	3	7	225	0,3312	0,5456	0,2137
0,15	3	10	150	0,3349	0,5361	0,2180
0,15	3	10	175	0,3329	0,5412	0,2159
0,15	3	10	200	0,3315	0,5450	0,2142
0,15	3	10	225	0,3302	0,5485	0,2127
0,15	3	12	150	0,3350	0,5358	0,2182
0,15	3	12	175	0,3330	0,5412	0,2159
0,15	3	12	200	0,3313	0,5456	0,2141
0,15	3	12	225	0,3298	0,5497	0,2125
0,15	3	15	150	0,3357	0,5342	0,2183
0,15	3	15	175	0,3337	0,5394	0,2163
0,15	3	15	200	0,3318	0,5443	0,2144
0,15	3	15	225	0,3306	0,5476	0,2129
0,15	4	7	150	0,3307	0,5470	0,2137
0,15	4	7	175	0,3291	0,5515	0,2117
0,15	4	7	200	0,3279	0,5547	0,2101
0,15	4	7	225	0,3268	0,5577	0,2084
0,15	4	10	150	0,3305	0,5475	0,2135
0,15	4	10	175	0,3290	0,5514	0,2115
0,15	4	10	200	0,3279	0,5544	0,2098
0,15	4	10	225	0,3272	0,5562	0,2085
0,15	4	12	150	0,3301	0,5487	0,2131
0,15	4	12	175	0,3285	0,5531	0,2111

Lernrate	Interaktionstiefe	n.minobsinnode	n.trees	RMSE	Rsquared	MAE
0,15	4	12	200	0,3272	0,5568	0,2094
0,15	4	12	225	0,3260	0,5599	0,2078
0,15	4	15	150	0,3298	0,5497	0,2132
0,15	4	15	175	0,3281	0,5541	0,2113
0,15	4	15	200	0,3266	0,5582	0,2096
0,15	4	15	225	0,3255	0,5612	0,2081
0,15	5	7	150	0,3279	0,5546	0,2099
0,15	5	7	175	0,3267	0,5579	0,2081
0,15	5	7	200	0,3255	0,5614	0,2064
0,15	5	7	225	0,3247	0,5636	0,2051
0,15	5	10	150	0,3264	0,5588	0,2096
0,15	5	10	175	0,3254	0,5616	0,2078
0,15	5	10	200	0,3243	0,5645	0,2061
0,15	5	10	225	0,3238	0,5661	0,2049
0,15	5	12	150	0,3274	0,5560	0,2100
0,15	5	12	175	0,3259	0,5601	0,2080
0,15	5	12	200	0,3247	0,5635	0,2064
0,15	5	12	225	0,3238	0,5658	0,2049
0,15	5	15	150	0,3264	0,5588	0,2096
0,15	5	15	175	0,3249	0,5626	0,2077
0,15	5	15	200	0,3241	0,5649	0,2064
0,15	5	15	225	0,3233	0,5671	0,2052
0,15	6	7	150	0,3262	0,5591	0,2074
0,15	6	7	175	0,3249	0,5627	0,2055
0,15	6	7	200	0,3242	0,5648	0,2040
0,15	6	7	225	0,3235	0,5667	0,2025
0,15	6	10	150	0,3249	0,5628	0,2070
0,15	6	10	175	0,3238	0,5656	0,2050
0,15	6	10	200	0,3229	0,5683	0,2032
0,15	6	10	225	0,3225	0,5695	0,2021
0,15	6	12	150	0,3245	0,5638	0,2066
0,15	6	12	175	0,3234	0,5670	0,2046
0,15	6	12	200	0,3225	0,5694	0,2030
0,15	6	12	225	0,3218	0,5713	0,2016
0,15	6	15	150	0,3255	0,5613	0,2074
0,15	6	15	175	0,3245	0,5640	0,2057
0,15	6	15	200	0,3236	0,5666	0,2043
0,15	6	15	225	0,3229	0,5686	0,2029
0,15	7	7	150	0,3248	0,5630	0,2053
0,15	7	7	175	0,3241	0,5653	0,2034
0,15	7	7	200	0,3236	0,5670	0,2018
0,15	7	7	225	0,3231	0,5684	0,2004
0,15	7	10	150	0,3235	0,5666	0,2046
0,15	7	10	175	0,3229	0,5683	0,2029
0,15	7	10	200	0,3227	0,5693	0,2015
0,15	7	10	225	0,3220	0,5714	0,2001
0,15	7	12	150	0,3241	0,5648	0,2051
0,15	7	12	175	0,3232	0,5676	0,2032
0,15	7	12	200	0,3228	0,5690	0,2017
0,15	7	12	225	0,3223	0,5703	0,2004
0,15	7	15	150	0,3226	0,5690	0,2048
0,15	7	15	175	0,3217	0,5717	0,2030

Lernrate	Interaktionstiefe	n.minobsinnode	n.trees	RMSE	Rsquared	MAE
0,15	7	15	200	0,3211	0,5732	0,2017
0,15	7	15	225	0,3208	0,5743	0,2004
0,15	8	7	150	0,3238	0,5659	0,2031
0,15	8	7	175	0,3231	0,5682	0,2012
0,15	8	7	200	0,3227	0,5696	0,1996
0,15	8	7	225	0,3221	0,5714	0,1981
0,15	8	10	150	0,3224	0,5696	0,2024
0,15	8	10	175	0,3221	0,5706	0,2008
0,15	8	10	200	0,3218	0,5720	0,1994
0,15	8	10	225	0,3216	0,5727	0,1981
0,15	8	12	150	0,3223	0,5698	0,2029
0,15	8	12	175	0,3217	0,5715	0,2010
0,15	8	12	200	0,3213	0,5728	0,1996
0,15	8	12	225	0,3210	0,5741	0,1983
0,15	8	15	150	0,3228	0,5687	0,2033
0,15	8	15	175	0,3218	0,5714	0,2015
0,15	8	15	200	0,3213	0,5731	0,1999
0,15	8	15	225	0,3210	0,5738	0,1987
0,15	9	7	150	0,3227	0,5694	0,2016
0,15	9	7	175	0,3220	0,5715	0,1996
0,15	9	7	200	0,3215	0,5731	0,1980
0,15	9	7	225	0,3214	0,5739	0,1967
0,15	9	10	150	0,3220	0,5711	0,2011
0,15	9	10	175	0,3216	0,5726	0,1995
0,15	9	10	200	0,3211	0,5741	0,1980
0,15	9	10	225	0,3210	0,5747	0,1969
0,15	9	12	150	0,3218	0,5714	0,2013
0,15	9	12	175	0,3214	0,5729	0,1995
0,15	9	12	200	0,3209	0,5744	0,1980
0,15	9	12	225	0,3209	0,5749	0,1967
0,15	9	15	150	0,3217	0,5716	0,2016
0,15	9	15	175	0,3211	0,5734	0,1998
0,15	9	15	200	0,3205	0,5752	0,1983
0,15	9	15	225	0,3202	0,5762	0,1971
0,2	3	7	150	0,3335	0,5394	0,2156
0,2	3	7	175	0,3318	0,5441	0,2138
0,2	3	7	200	0,3305	0,5476	0,2122
0,2	3	7	225	0,3298	0,5495	0,2109
0,2	3	10	150	0,3332	0,5401	0,2153
0,2	3	10	175	0,3317	0,5444	0,2134
0,2	3	10	200	0,3303	0,5483	0,2119
0,2	3	10	225	0,3292	0,5511	0,2105
0,2	3	12	150	0,3335	0,5393	0,2160
0,2	3	12	175	0,3316	0,5446	0,2139
0,2	3	12	200	0,3303	0,5483	0,2123
0,2	3	12	225	0,3293	0,5507	0,2111
0,2	3	15	150	0,3332	0,5401	0,2156
0,2	3	15	175	0,3316	0,5446	0,2138
0,2	3	15	200	0,3303	0,5482	0,2122
0,2	3	15	225	0,3292	0,5513	0,2109
0,2	4	7	150	0,3307	0,5472	0,2118
0,2	4	7	175	0,3295	0,5506	0,2100

Lernrate	Interaktionstiefe	n.minobsinnode	n.trees	RMSE	Rsquared	MAE
0,2	4	7	200	0,3283	0,5539	0,2084
0,2	4	7	225	0,3278	0,5558	0,2070
0,2	4	10	150	0,3294	0,5507	0,2110
0,2	4	10	175	0,3281	0,5544	0,2091
0,2	4	10	200	0,3271	0,5573	0,2077
0,2	4	10	225	0,3265	0,5590	0,2063
0,2	4	12	150	0,3287	0,5523	0,2107
0,2	4	12	175	0,3273	0,5563	0,2088
0,2	4	12	200	0,3266	0,5586	0,2076
0,2	4	12	225	0,3254	0,5617	0,2060
0,2	4	15	150	0,3299	0,5495	0,2116
0,2	4	15	175	0,3280	0,5548	0,2097
0,2	4	15	200	0,3271	0,5573	0,2082
0,2	4	15	225	0,3261	0,5600	0,2069
0,2	5	7	150	0,3277	0,5551	0,2083
0,2	5	7	175	0,3269	0,5578	0,2063
0,2	5	7	200	0,3264	0,5595	0,2049
0,2	5	7	225	0,3260	0,5608	0,2037
0,2	5	10	150	0,3269	0,5577	0,2081
0,2	5	10	175	0,3260	0,5603	0,2064
0,2	5	10	200	0,3254	0,5621	0,2049
0,2	5	10	225	0,3249	0,5640	0,2036
0,2	5	12	150	0,3264	0,5591	0,2076
0,2	5	12	175	0,3254	0,5620	0,2059
0,2	5	12	200	0,3248	0,5639	0,2046
0,2	5	12	225	0,3242	0,5658	0,2034
0,2	5	15	150	0,3267	0,5581	0,2083
0,2	5	15	175	0,3258	0,5609	0,2066
0,2	5	15	200	0,3249	0,5634	0,2051
0,2	5	15	225	0,3244	0,5650	0,2038
0,2	6	7	150	0,3269	0,5583	0,2060
0,2	6	7	175	0,3261	0,5609	0,2044
0,2	6	7	200	0,3252	0,5635	0,2025
0,2	6	7	225	0,3251	0,5641	0,2014
0,2	6	10	150	0,3249	0,5633	0,2053
0,2	6	10	175	0,3240	0,5662	0,2034
0,2	6	10	200	0,3232	0,5683	0,2018
0,2	6	10	225	0,3228	0,5696	0,2005
0,2	6	12	150	0,3258	0,5610	0,2054
0,2	6	12	175	0,3248	0,5640	0,2037
0,2	6	12	200	0,3242	0,5658	0,2022
0,2	6	12	225	0,3239	0,5668	0,2012
0,2	6	15	150	0,3246	0,5640	0,2053
0,2	6	15	175	0,3239	0,5661	0,2036
0,2	6	15	200	0,3233	0,5681	0,2023
0,2	6	15	225	0,3232	0,5685	0,2012
0,2	7	7	150	0,3256	0,5617	0,2037
0,2	7	7	175	0,3249	0,5643	0,2017
0,2	7	7	200	0,3245	0,5655	0,2001
0,2	7	7	225	0,3244	0,5661	0,1990
0,2	7	10	150	0,3246	0,5645	0,2030
0,2	7	10	175	0,3241	0,5663	0,2014

Lernrate	Interaktionstiefe	n.minobsinnode	n.trees	RMSE	Rsquared	MAE
0,2	7	10	200	0,3236	0,5677	0,2001
0,2	7	10	225	0,3236	0,5684	0,1990
0,2	7	12	150	0,3241	0,5657	0,2033
0,2	7	12	175	0,3235	0,5674	0,2016
0,2	7	12	200	0,3232	0,5686	0,2001
0,2	7	12	225	0,3231	0,5695	0,1990
0,2	7	15	150	0,3247	0,5642	0,2038
0,2	7	15	175	0,3241	0,5661	0,2021
0,2	7	15	200	0,3239	0,5671	0,2008
0,2	7	15	225	0,3238	0,5677	0,1996
0,2	8	7	150	0,3259	0,5616	0,2023
0,2	8	7	175	0,3256	0,5626	0,2007
0,2	8	7	200	0,3255	0,5638	0,1992
0,2	8	7	225	0,3253	0,5647	0,1979
0,2	8	10	150	0,3246	0,5647	0,2016
0,2	8	10	175	0,3239	0,5669	0,2000
0,2	8	10	200	0,3238	0,5678	0,1987
0,2	8	10	225	0,3239	0,5680	0,1976
0,2	8	12	150	0,3231	0,5687	0,2012
0,2	8	12	175	0,3230	0,5694	0,1997
0,2	8	12	200	0,3227	0,5704	0,1983
0,2	8	12	225	0,3228	0,5707	0,1972
0,2	8	15	150	0,3239	0,5664	0,2020
0,2	8	15	175	0,3237	0,5674	0,2005
0,2	8	15	200	0,3233	0,5688	0,1991
0,2	8	15	225	0,3231	0,5697	0,1980
0,2	9	7	150	0,3246	0,5652	0,2004
0,2	9	7	175	0,3247	0,5656	0,1988
0,2	9	7	200	0,3247	0,5663	0,1974
0,2	9	7	225	0,3249	0,5664	0,1964
0,2	9	10	150	0,3244	0,5659	0,1999
0,2	9	10	175	0,3240	0,5673	0,1984
0,2	9	10	200	0,3242	0,5673	0,1972
0,2	9	10	225	0,3244	0,5674	0,1963
0,2	9	12	150	0,3227	0,5700	0,1997
0,2	9	12	175	0,3226	0,5706	0,1983
0,2	9	12	200	0,3226	0,5712	0,1970
0,2	9	12	225	0,3228	0,5712	0,1959
0,2	9	15	150	0,3236	0,5676	0,2003
0,2	9	15	175	0,3233	0,5690	0,1988
0,2	9	15	200	0,3229	0,5703	0,1975
0,2	9	15	225	0,3229	0,5706	0,1964

Stochastic Gradient Boosting
 6238 Stichproben
 20 Prädiktoren
 keine Vorverarbeitung
 Resampling: Kreuzvalidierung (10-fach, 10 Wiederholungen)
 Stichprobengrößen: 5613, 5613, 5615, 5615, 5613, 5615, ...

Tabelle 11: RMSE der jeweils besten Modelle der verschiedenen Ansätze (gerundet)

Modell	RMSE
RF_fein	0,6003
RF	0,5993
GBM_fein	0,3196
GBM	0,3398
SVM	0,4992
EN	0,4781

Tabelle 12: Wichtigkeiten der UV im finalen RF-Modell (gerundet)

Variable	%IncMSE	IncNodePurity
n1_add	30,5011	37,0971
n2_add	21,2987	17,9909
n3_add	48,1588	44,4813
n4_add	38,4314	66,9131
n5_add	82,7338	243,0454
t1_add	15,5552	14,6716
t2_add	15,0966	10,6068
t3_add	18,5161	17,2172
t4_add	24,7408	25,5602
t5_add	47,4267	106,6275
n1_del	49,9704	157,5354
n2_del	26,1669	38,9715
n3_del	26,0787	23,0193
n4_del	28,7627	46,4524
n5_del	24,8936	76,8863
t1_del	18,6448	39,2200
t2_del	16,7577	8,6760
t3_del	15,2034	9,2172
t4_del	23,0260	27,9505
t5_del	22,0310	33,0532

Literaturverzeichnis

- Akerlof, G. A. (1970). The Market for “Lemons”: Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics* 84(3), 488-500
- Alpaydin, E. (2008). *Maschinelles Lernen*. München: Oldenbourg.
- Assenmacher, W. (2000). *Induktive Statistik*. Berlin et al.: Springer.
- Ba, S. & Pavlou, P. A. (2002). Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *MIS Quarterly* 26(3), 243-268
- Bakos, Y (1997). Reducing Buyer Search Costs: Implications for Electronic Marketplaces. *Management Science* 43(12), 1676-1692
- Beel, J. & Gipp, B. (2009). Google Scholar’s Ranking Algorithm: An Introductory Overview. *Proceedings of 3rd International Conference on Research Challenges in Information Science*. Institute of Electrical and Electronics Engineers. (<https://gipp.com/wp-content/papercite-data/pdf/beel09.pdf>, Zugriff: 20.03.2019)
- Bolton, G. E., Katok, E. & Ockenfels, A. (2004). How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation. *Management Science* 50(11), 1587-1602
- Burger, S. V. (2018). *Introduction to Machine Learning with R: Rigorous Mathematical Modeling*. Beijing, Boston: O’Reilly.
- Chen, L., Mislove, A. & Wilson, C. (2016). An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace. *Proceedings of the 25th International Conference on World Wide Web*. Canada: ACM. (<https://doi.org/10.1145/2872427.2883089>, Zugriff: 20.03.2019)
- Chen, P.-Y., Dhanasobhon, S. & Smith, M. D. (2008). All Reviews are Not Created Equal: The Disaggregate Impact of Reviews and Reviewers at Amazon.com. *Arbeitspapier, Carnegie Mellon University* (<http://ssrn.com/abstract=918083>, Zugriff: 19.03.2019)
- Chen, P.-Y., Wu, S.-Y. & Yoon, J. (2004). The Impact of Online Recommendations and Consumer Feedback on Sales. *ICIS 2004 Proceedings* 58, 711-723
- Chevalier, J. A. & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research* 43(3), 345-354
- Coster, H. (2006). The Secret Life of an Online Book Reviewer. *Forbes Magazine*, 01. Dezember 2006 (https://www.forbes.com/2006/11/30/donald-mitchell-amazon-tech-media_cz_hc_books06_1201amazon.html#cd168e578bc6, Zugriff: 20.03.2019)
- Dabholkar, P. A. (2006). Factors Influencing Consumer Choice of a “Rating Web Site”: An Experimental Investigation of an Online Interactive Decision Aid. *Journal of Marketing Theory and Practice* 14(4), 259-273
- Dellarocas, C. (2003). The Digitalization of Word of Mouth: Promise Challenges of Online Feedback Mechanisms. *Management Science* 49(10), 1407-1424

- Djawadi, B. M., Fahr, R., Haake, C.-J. & Recker, S. (2018). Maintaining vs. Milking Good Reputation When Customer Feedback Is Inaccurate. *Public Library of Science* 13(11), e0207172 (<https://doi.org/10.1371/journal.pone.0207172>, Zugriff: 20.03.2019)
- Duan, W., Gu, B. & Whinston, A. B. (2008). Do Online Reviews Matter? – An Empirical Investigation of Panel Data. *Decision Support Systems* 45, 1007-1016
- Fang, G.-S., Kamei, S. & Fujita, S. (2017). Rating Prediction with Topic Gradient Descent Method for Matrix Factorization in Recommendation. *International Journal of Advanced Computer Science and Applications* 8(12), 469-476
- Forman, C., Ghose, A. & Wiesenfeld, B. (2008). Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research* 19(3), 291-313
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 29(5), 1189-1232
- Fung, R. & Lee, M. (1999). EC-Trust (Trust in Electronic Commerce): Exploring the Antecedent Factors. In: Hasemann, W. D. & Nazareth, D. L. (Herausgeber). *Proceedings of the Fifth Americas Conference on Information Systems*, 517-519. Association for Information Systems.
- Gretzel, U. & Fesenmaier, D. R. (2006). Persuasion in Recommender Systems. *International Journal of Electronic Commerce* 11(2), 81-100
- Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. (2. Auflage). New York: Springer.
- Hawes, J. M., Mast, K. E. & Swan, J. E. (1989). Trust Earning Perceptions of Sellers and Buyers. *Journal of Personal Selling & Sales Management* 9(1), 1-8
- Hübler, O. (1989). *Ökonometrie*. Stuttgart, New York: Gustav Fisher Verlag.
- Jin, G. Z. & Kato, A. (2006). Price, Quality, and Reputation: Evidence from an Online Field Experiment. *The RAND Journal of Economics* 37(4), 983-1004
- Kaplan, K. J. (1972). On the Ambivalence-Indifference Problem in Attitude Theory Measurement: A Suggested Modification of the Semantic Differential Technique. *Psychological Bulletin* 77(5), 361-372
- Kelleher, J. D., Mac Namee, B. & D'Arcy, A. (2015). *Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. Cambridge, Massachusetts: The MIT Press.
- Knox, S. W. (2018). *Machine Learning: A Concise Introduction*. Hoboken: Wiley.
- Kumar, N. & Benbasat, I. (2006). Research Note: The Influence of Recommendations and Customer Reviews on Evaluations of Websites. *Information Systems Research* 17(4), 425-439
- Miller, N., Resnick, P. & Zeckhauser, R. (2005). Eliciting Informative Feedback: The Peer-Prediction Method. *Management Science* 51(9), 1359-1373

- Mohanty, N., Lee-St. John, A., Manmatha, R. & Rath, T. M. (2013). In Govindaraju, V. (Herausgeber). *Machine Learning: Theory and Applications*, 249-267. Amsterdam et al.: Elsevier.
- Mudambi, S. M. & Schuff, D. (2010). Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly* 34(1), 185-200
- Ngo-Ye, T. L. & Sinha, A. P. (2014). The Influence of Reviewer Engagement Characteristics on Online Review Helpfulness: A Text Regression Model. *Decision Support Systems* 61, 47-58
- Pavlou, P. A. & Dimoka, A. (2006). The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation. *Information Systems Research* 17(4), 392-414
- Peters, R. & Reitzenstein, I. (2008). Reputationssysteme im eCommerce – Funktionsweise, Anwendung und Nutzenpotenziale. *HMD Praxis der Wirtschaftsinformatik* 45(3), 43-50
- Presser, S. & Schuman, H. (1980). The Measurement of a Middle Position in Attitude Surveys. *Public Opinion Quarterly* 44(1), 70-85
- Reitzenstein, I. & Peters, R. (2009). Assessing Robustness of Reputation Systems Regarding Interdependent Manipulations, In: Di Noia, T. & Buccafurri, F. (Herausgeber). *E-Commerce and Web Technologies. EC-Web 2009. Lecture Notes in Computer Science* 5692, 288-299. Berlin, Heidelberg: Springer.
- Resnick, P. & Zeckhauser, R. (2015). Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. *The Economics of the Internet and E-commerce* 11, 127-157
- Resnick, P., Zeckhauser, R., Friedman, E. & Kuwabara, K. (2000). Reputation Systems. *Communication of the ACM* 43(12), 45-48
- Shen, W., Hu, Y. J. & Ulmer, J. R. (2015). Competing for Attention: An Empirical Study of Online Reviewers' Strategic Behavior. *MIS Quarterly* 39(3), 683-696
- Smith, M. D. & Brynjolfsson, E. (2001). Consumer Decision-Making at an Internet Shopbot: Brand Still Matters. *Industrial Economics* 49(4), 541-558
- Stigler, G. J. (1961). The Economics of Information. *Journal of Political Economy* 69(3), 213-225
- Swan, J. E. & Nolan, J. J. (1985). Gaining Customer Trust: A Conceptual Guide for the Salesperson. *Journal of Personal Selling & Sales Management* 5(2), 39-48
- Walden, E. (2000). Some Value Propositions of Online Communities. *Electronic Markets* 10(15), 244-249
- Wan, Y & Nakayama, M. (2014). The Reliability of Online Review Helpfulness. *Journal of Electronic Commerce Research* 15(3), 179-189
- Zacharia, G., Moukas, A. & Maes, P. (2000). Collaborative Reputation Mechanisms for Electronic Marketplaces. *Decision Support Systems* 29, 371-388

- Zhou, Y.-B., Lei, T & Zhou, T. (2011). A Robust Ranking Algorithm to Spamming. *Europhysics Letters* 94, 48002
- Ziegele, M. & Weber, M. (2015). Example, Please! Comparing the Effects of Single Customer Reviews and Aggregate Review Scores on Online Shoppers' Product Evaluations. *Journal of Consumer Behaviour* 14, 103-114

Internetquellenverzeichnis

- Amalytix.com [1]: <https://www.amalytix.com/impressum> (Zugriff: 20.03.2019)
- Amalytix.com [2]: <https://www.amalytix.com/funktionen> (Zugriff: 20.03.2019)
- Amazon.de [1]: <https://www.amazon.de/gp/forum/content/qa-guidelines.html> (Zugriff: 20.03.2019)
- Amazon.de [2]: <https://www.amazon.de/gp/help/customer/display.html?nodeId=201470700> (Zugriff: 20.03.2019)
- Amazon.de [3]: <https://www.amazon.de/gp/help/customer/display.html?nodeId=201929730> (Zugriff: 20.03.2019)
- Amazon.de [4]: https://www.amazon.de/Schmidt-Spiele-49020-Jubil%C3%A4umsausgabe-Holzsteinen/dp/B00006RTA3/ref=sr_1_1?__mk_de_DE=%C3%85M%C3%85%C5%BD%C3%95%C3%91&crid=18A241M2D007K&key-words=mensch+%C3%A4rgere+dich+nicht&qid=1553087078&s=gateway&sprefix=mensch+%2Caps%2C158&sr=8-1 (Zugriff: 19.03.2019)
- Statista.com [1]: <https://de.statista.com/statistik/daten/studie/860277/umfrage/top-online-shops-weltweit-ecommercedb> (Zugriff: 20.03.2019)

Ehrenwörtliche Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel verfasst habe. Wörtlich übernommene Sätze oder Satzteile sind als Zitat belegt, andere Anlehnungen, hinsichtlich Aussage und Umfang, unter Quellenangabe kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen und ist nicht veröffentlicht. Sie wurde nicht, auch nicht auszugsweise, für eine andere Prüfungs- oder Studienleistung verwendet. Zudem versichere ich, dass die von mir abgegebenen schriftlichen (gebundenen) Versionen der vorliegenden Arbeit mit der abgegebenen elektronischen Version auf einem Datenträger inhaltlich übereinstimmen.

Ort, Datum: _____ Unterschrift: _____