

exp



universität
uulm

Universität Ulm
Fakultät für Mathematik und
Wirtschaftswissenschaften

Data Analytics in der Tourismusbranche am Beispiel der Firma Wilken

Masterarbeit

In Wirtschaftsmathematik

vorgelegt von
Oechslein, Felix, 1010801, Wirtschaftsmathematik, Semester 11
am 31.05.2023

Gutachter

Prof. Dr. Mathias Klier
Prof. Dr. Mitja Stadje

Sperrvermerk

Diese Arbeit enthält Informationen und Daten, welche die Belange und Interessen der Firma Wilken betreffen. Angegebene Daten und Informationen dürfen nicht an Dritte weitergegeben oder in irgendeiner Form veröffentlicht bzw. kopiert werden, auch nicht auszugsweise.

Ort, Datum

Vorname Nachname

Inhaltsverzeichnis

Sperrvermerk.....	II
Abbildungsverzeichnis	V
Tabellenverzeichnis	VI
1 Einleitung.....	1
2 Problemstellung	4
3 Kundenbindung und E-Mail-Marketing	5
3.1 Kundenbindung im Kulturbereich	5
3.2 Das Thalia Theater	8
3.3 E-Mail-Marketing des Thalia Theaters	11
3.4 Öffnungserkennung	12
3.5 Nutzen der E-Marketing Suite der Firma Wilken	13
4 Das Hidden Markov Model.....	15
4.1 Intuition.....	15
4.2 Elemente eines Hidden Markov Models.....	18
5 Beschreibung des verwendeten Modells.....	20
5.1 Startverteilung	20
5.2 Übergangsmatrix	20
5.3 Bedingte E-Mail-Öffnungsanzahl (CEOM)	21
5.4 Modell der Kaufanzahl (CPM).....	22
5.5 Die Korrelation zwischen Kaufentscheidung und E-Mail-Öffnungen.....	24
5.6 Modellbestimmung.....	25
6 Auswertung.....	26
6.1 Datenauswahl.....	26
6.2 Modellidentifikation	27
6.3 Parameterbestimmungen.....	31
6.4 Bestimmung des Kundenzustands.....	37

Abbildungsverzeichnis	IV
6.5 Zusammenfassung	41
6.6 Robustheitstests	41
7 Schlussfolgerungen, Beschränkungen und zukünftige Forschung	45
Anlage	Fehler! Textmarke nicht definiert.
Literaturverzeichnis.....	VII
Ehrenwörtliche Erklärung.....	VIII

Abbildungsverzeichnis

Abbildung 1: Konzeptualisierung der Kundenbindung (Bruhn 2005), S. 8	6
Abbildung 2: Anzahl der Unternehmen im Markt für darstellende Künste in Deutschland von 2003 bis 2021 (Statista 2023)	7
Abbildung 3: Ticketverkäufe des Thalia Theaters in den Jahren 2018 bis 2022 nach Quartalen	9
Abbildung 4: Aggregierte Verkäufe pro Monat (2014 bis 2019)	10
Abbildung 5: Versendete E-Mails im Zeitraum Oktober 2021 bis Dezember 2022	11
Abbildung 6: Einbindung der E-Marketing Suite in die CRM-Landschaft des Theater Basel (Franziska Pinkert et al. 2022)	13
Abbildung 7: Modellierung von verborgenen Münzwurf-Experimenten (Rabiner und Juang 1986, S. 6)	
Abbildung 8: Histogramm der Ticketverkäufe (Januar 2021 bis Dezember 2022)	26
Abbildung 9: Histogramm der Käufe pro Monat und Kunde (im Zeitraum Oktober 2021 bis Dezember 2022)	28
Abbildung 10: Vergleich der Verteilungsfunktionen (CDF) einer ZINBD-, NBD- und Poissonverteilten ZV mit geschätzten Parameterwerten gegenüber der empirischen Verteilungsfunktion	30
Abbildung 11: Wahrscheinlichkeitsfunktion der Käufe für verschiedene Werte von λ ..	33
Abbildung 13: Vergleich der Poissonverteilung, NBD und der empirischen Verteilung der Öffnungszahlen.....	44

Tabellenverzeichnis

Tabelle 1: Startparameter und Grenzen (für die einzelnen Werte in den Parametern)	31
Tabelle 3: Relevante Parameter für die Modellierung des Kaufverhaltens	32
Tabelle 4: Einfluss der E-Mail-Kampagnen auf das Kaufverhalten	33
Tabelle 5: Relevante Parameter für die Modellierung der E-Mail-Öffnungen	34
Tabelle 6: Veranschaulichung der Übergangsmatrix anhand eines „typischen“ Kunden (eine Öffnung im letzten Monat und kein Kauf)	35
Tabelle 7: Veranschaulichung der Übergangsmatrix anhand eines Kunden mit einer Öffnung im letzten Monat und einem Kauf	35
Tabelle 8: Veranschaulichung der Übergangsmatrix anhand eines Kunden mit keiner Öffnung im letzten Monat und keinem Kauf	36

1 Einleitung

Diese Arbeit findet ihren Ursprung in einem Gespräch des Abteilungsleiters Tourismus der Firma Wilken mit dem Marketingleiter eines der größten deutschen Theater. Darin wurde die Frage aufgeworfen:

„Ist die Wirkung des E-Mail-Marketings auf die langfristige Kundenbindung denn messbar und wird mit dem E-Mail-Marketing ein positiver wirtschaftlicher Effekt erzielt?“

Da die Kundenbindung eine komplexe Beziehung zwischen Kunden und Unternehmen darstellt, ist diese Frage grundsätzlich schwer zu beantworten. Dennoch gibt es Möglichkeiten, Veränderungen im Verhalten von Kunden nach verschiedenen Interaktionen mit dem Unternehmen zu untersuchen und daraus Aufschlüsse über die zugrundeliegende Beziehung zwischen Kunden und Unternehmen zu gewinnen.

E-Mail-Kampagnen spielen bei der Stärkung der Kundenbindung eine zentrale Rolle, da dadurch ein regelmäßiger Kontakt zwischen Unternehmen und Kunden aufgebaut werden kann. Dadurch, dass sowohl die Kundenbindung, als auch der aus einer Stärkung der Kundenbindung resultierende Gewinn nur schwer messbar sind, sehen sich Unternehmen bei der effizienten Durchführung von E-Mail-Kampagnen mit mehreren Herausforderungen konfrontiert (vgl. Zhang et al. 2017, S. 851f.).

Erstens besteht die Problematik, dass Kunden, die regelmäßig E-Mails eines Unternehmens öffnen, nicht zwangsläufig auch viele Käufe bei diesem Unternehmen tätigen, und umgekehrt. Beispielsweise können Stammkunden bereits mit dem Unternehmen und dessen Angeboten bestens vertraut sein und daher weniger Anreiz haben, die E-Mails des Unternehmens regelmäßig zu öffnen. Daher ist es für das Unternehmen möglicherweise nicht erforderlich, solche Kunden intensiv durch E-Mail-Marketing anzusprechen, da diese Kategorie von Kunden den empfangenen E-Mails wenig bis gar keine Beachtung schenkt. Andererseits gibt es Kunden mit geringeren Kaufaktivitäten, die jedoch besonders auf das Öffnen von E-Mails des Unternehmens ansprechen. Diese Kunden suchen gezielt nach Informationen in den E-Mails. Trotz der höheren Öffnungsraten von E-Mails bei dieser Kundengruppe ist es aus finanzieller Sicht möglicherweise nicht optimal, diese Kunden verstärkt anzusprechen, da die Wahrscheinlichkeit einer wiederholten Kaufaktivität eher gering ist.

Zweitens ist das Versenden der richtigen Anzahl von E-Mails für die Rentabilität des Unternehmens von entscheidender Bedeutung, insbesondere da die meisten Kunden sich über die große Anzahl von E-Mails, die Unternehmen versenden, beschweren. Die Studie „E-Mail-Marketing-Benchmark“ der Firma Inxmail GmbH (2023) zeigt auf, dass die Öffnungsrate der E-Mail-Kampagnen deutscher Unternehmen von 30.6% im Jahr 2021 auf 23.9% im Jahr 2022

gesunken ist. Unternehmen versuchen diesem Umstand entgegenzuwirken, indem sie die Versandfrequenz erhöhen. Um negative Auswirkungen von zu vielen E-Mails zu vermeiden, empfiehlt Inxmail eine Anpassung der Versandfrequenz an die jeweiligen Vorlieben der Kunden. Obwohl es von größter Bedeutung ist, die optimale Anzahl von E-Mails zu versenden, gestaltet sich die Suche nach dieser "magischen" Zahl für ein Unternehmen äußerst herausfordernd. Dies liegt daran, dass nicht nur die Kunden unterschiedliche intrinsische Präferenzen für E-Mails haben (d.h. die Öffnungshäufigkeit von E-Mails variiert zwischen den Kunden), sondern auch die Kundenpräferenzen sich im Laufe der Zeit dynamisch verändern können (d.h. die E-Mail-Öffnungshäufigkeit eines individuellen Kunden kann im Laufe der Zeit variieren). Mit anderen Worten: Die optimale Anzahl von E-Mails, die versendet werden sollten, kann sich von Kunde zu Kunde und im Laufe der Zeit unterscheiden.

Drittens neigen Führungskräfte dazu, jede E-Mail-Marketing-Kampagne als eigenständigen Akquisitionsprozess zu betrachten und vernachlässigen dabei häufig den langfristigen Einfluss auf das Öffnungsverhalten von E-Mails und das Kaufverhalten der Kunden. Neben den unmittelbaren Effekten kann eine E-Mail auch langfristige Auswirkungen auf die Rentabilität des Unternehmens haben, indem sie das Kunden-Firmen-Beziehungsverhältnis beeinflusst. Kunden, die E-Mails öffnen, können ein gesteigertes Interesse an dem Unternehmen und seinen Angeboten entwickeln, was wiederum zu einer erhöhten Kaufbereitschaft führen kann. Die Nichtbeachtung der langfristigen Auswirkungen von E-Mails könnte daher dazu führen, dass das Unternehmen suboptimale Entscheidungen im Bereich des E-Mail-Marketings trifft.

Aufgrund all dieser Herausforderungen benötigen Unternehmen eine Orientierungshilfe in Form einer Entscheidungsunterstützung, um die optimale Anzahl von E-Mails zu bestimmen, die im Laufe der Zeit an verschiedene Kunden gesendet werden sollen. In dieser Arbeit ist das Ziel, eine Entscheidungshilfe für Unternehmen bereitzustellen, indem erstens das Öffnungsverhalten von Kunden bei E-Mails mit ihrem Kaufverhalten verknüpft wird und zweitens die langfristigen Auswirkungen des E-Mail-Marketings auf die Rentabilität des Unternehmens berücksichtigt wird.

Methodisch orientiert sich diese Arbeit an Zhang et al. (2017). Das verwendete Modell besteht aus einem Hidden Markow Modell (HMM) zur Erfassung der unterschiedlichen Arten der Kundenbindung und des Verhaltens der Kunden mit einer ähnlichen Kundenbindung. Des Weiteren wird ein Copula-Modell zur Modellierung der Abhängigkeiten des Kauf- und E-Mail-Verhaltens im Zeitverlauf angewendet. Die Kombination der beiden Modelle wird dazu genutzt, Kunden in unterschiedliche Segmente zu unterteilen, wobei Kunden in den einzelnen Segmenten ein ähnliches Kauf- und E-Mail-Verhalten aufweisen. Basierend auf der Segmentierung der Kunden, wird der Einfluss des E-Mail-Versands auf ihre Bereitschaft in ein anderes Segment

zu wechseln und der Stärkung oder Schwächung ihres Kaufverhaltens untersucht. Die Kombination aus Segmentierung der Kunden und Untersuchung des unterschiedlichen Einflusses der Häufigkeit des E-Mail-Versands auf das Kaufverhalten kann genutzt werden, um die E-Mail-Frequenz an die Kundengruppen anzupassen und dadurch die langfristige Kundenbindung zu erhöhen.

2 Problemstellung

Grundsätzlich wird davon ausgegangen, dass die tatsächliche Kundenbindung nicht mit Transaktionsdaten und E-Mail-Daten messbar ist und das Verhalten der Kunden in diesen Bereichen lediglich ein Indiz für die tatsächliche Kundenbindung darstellt. Es wird ein HMM verwendet, um die Kundenbindungsniveaus als versteckte Ausgangsgrößen zu bestimmen, die ein bestimmtes Verhalten wahrscheinlicher machen. Über die Bestimmung von Ähnlichkeiten im Verhalten der Kunden wird eine Segmentierung der Kunden erzielt. Die Segmente der einzelnen Kundengruppen haben wiederum ein ähnliches Kaufverhalten bei dem Erhalt einer ähnlichen Anzahl an E-Mails. Im Vergleich zu anderen Studien im Bereich der HMM, die sich nur auf das Kaufverhalten des Kunden konzentrieren und die Möglichkeit ignorieren, dass andere nicht-kaufbezogene Aktivitäten ebenfalls die Kunden-Firmen-Beziehung beeinflussen können, unterscheidet sich das Modell von Zhang et al. (2017) darin, dass auch nicht-kaufbezogene Aktivitäten wie die Öffnung einer E-Mail bei der Bestimmung der zugrundeliegenden Kunden-Firmen-Beziehung berücksichtigt werden.

Das in dieser Arbeit betrachtete Unternehmen ist das Thalia Theater in Hamburg. Als Staatstheater ist es finanziell abhängig von dem Haushalt der Stadt Hamburg. Dadurch dass ein festgesetztes Budget für den Betrieb des Unternehmens vorgegeben wird, spielt insbesondere eine gute Kostenpolitik eine wichtige Rolle bei wirtschaftlichen Erwägungen. Somit ist eine richtige Allokation von Marketingressourcen sehr wichtig für das Unternehmen. Zudem steht das Thalia Theater wie andere Anbieter für die Freizeitgestaltung vor der Herausforderung eines immer größer werdenden Angebots anderer Dienstleister. Dadurch nimmt die Kundenbindung eine wichtige Rolle zum Erhalt von Marktanteilen ein. Das in dieser Arbeit vorgestellte Modell kann dabei helfen, die richtigen Kunden anzusprechen und bei der richtigen Allokation von Marketingressourcen unterstützen. Für die Analyse der Kunden des Thalia Theaters und der Schätzung der Modellparameter des HMM stehen Daten zur Öffnung von E-Mail-Kampagnen und zum Kauf von Theatertickets zur Verfügung.

3 Kundenbindung und E-Mail-Marketing

In diesem Kapitel werden die theoretischen Grundlagen für die Arbeit dargestellt. Zunächst wird die Kundenbindung im Kulturbereich definiert und veranschaulicht, dass die Kundenbindung auch nicht-kaufbezogene Verhaltensweisen berücksichtigt. Es wird die Wirkungskette der Kundenbindung dargestellt und der Einfluss des E-Mail-Marketings auf diese Wirkungskette beschrieben. Anschließend wird das Thalia Theater, dessen wirtschaftliche Situation und dessen Strategie für das E-Mail-Marketing genauer analysiert. Das Kapitel schließt mit der Diskussion der Zuverlässigkeit von Öffnungserkennungen im Zeitalter von Mail Privacy Protection Software ab.

3.1 Kundenbindung im Kulturbereich

Im Theatermarkt in Hamburg spielt die Kundenbindung eine zentrale Rolle. Theaterhäuser konkurrieren um die Aufmerksamkeit und Treue der Zuschauer, da diese maßgeblich zum Erfolg und zur langfristigen Stabilität des Theaters beitragen. Kundenbindung bezieht sich auf die Bemühungen eines Theaters, eine starke und langfristige Beziehung zu seinen Zuschauern aufzubauen. Hamburg ist eine Stadt mit einer reichen Theaterlandschaft, die eine Vielzahl von Genres und Aufführungen bietet. In einer Zeit, in der die Unterhaltungsindustrie viele Alternativen bietet, ist es für Theaterhäuser von entscheidender Bedeutung, ihre Kunden zu binden und sicherzustellen, dass sie regelmäßig wiederkommen. Eine effektive Kundenbindung erfordert auch eine gezielte Kommunikation und Interaktion mit dem Publikum. Durch personalisierte Werbung und regelmäßige Newsletter können Theaterhäuser eine Beziehung zu ihren Zuschauern aufbauen und sie über kommende Aufführungen, Sonderveranstaltungen oder Rabattaktionen informieren. Im Folgenden werden die strategische Bedeutung und der Begriff der Kundenbindung erläutert. Danach wird die Wirkungskette der Kundenbindung untersucht.

Definition der Kundenbindung

Der Begriff der Kundenbindung wird in der Literatur auf unterschiedliche Weise verwendet. Es gibt verschiedene Definitionen, die unter anderem darauf zurückzuführen sind, dass Kundenbindung aus verschiedenen Perspektiven betrachtet werden kann. Dabei spielen die Sichtweise des Anbieters, die Sichtweise des Kunden und die Betrachtung der Geschäftsbeziehung eine Rolle. Wir beschränken uns in dieser Arbeit auf die Definition von Bruhn 2005, S. 7:

„Kundenbindung umfasst sämtliche Maßnahmen eines Unternehmens, die darauf abzielen, sowohl die Verhaltensabsichten als auch das tatsächliche Verhalten eines Kunden gegenüber einem Anbieter oder dessen Leistungen positiv zu gestalten, um die Beziehung zu diesem Kunden für die Zukunft zu stabilisieren bzw. auszuweiten.“

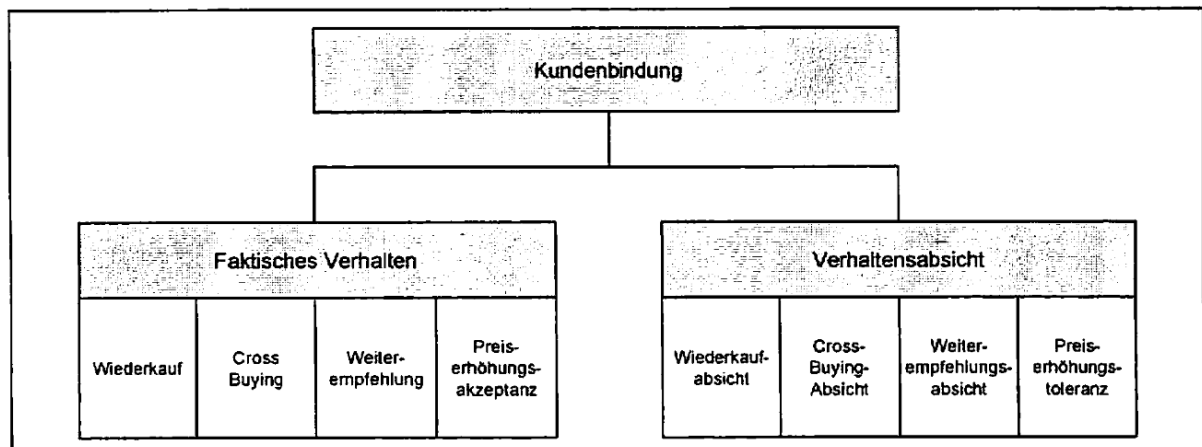


Abbildung 1: Konzeptualisierung der Kundenbindung (Bruhn 2005), S. 8

Diese Definition macht deutlich, dass die Messung der Stärke einer Kundenbindung sowohl das tatsächliche Verhalten als auch die Absicht des Verhaltens berücksichtigen muss. Hieraus leitet sich das Konstrukt der Kundenbindung, wie in **Fehler! Verweisquelle konnte nicht gefunden werden.** dargestellt, ab.

Strategische Bedeutung der Kundenbindung

Im Kulturbereich zeigt sich eine steigende Anspruchshaltung und eine zunehmende Unbeständigkeit der Konsumenten, die individuellere Bedürfnisse und ein breiteres Verständnis ihres Konsums aufweisen. Die Konkurrenz beschränkt sich nicht mehr nur auf andere Kulturangebote, sondern erstreckt sich auf die gesamte Freizeit-, Unterhaltungs- und Sinnstiftungsindustrie sowie die sozialen Netzwerke. Vor diesem Hintergrund gewinnen die Individualisierung und die Kundenbindung zunehmend an Bedeutung. Die Anbieter reagieren darauf mit maßgeschneiderten Angeboten, individueller Kundenansprache und einer persönlichen Rundum-Betreuung. Ähnlich wie Anlageberater in Banken, wird der Kulturkonsumentenberater zu einer neuen Dienstleistung in der Zielgruppenpflege (vgl. swissfuture 2014, S. 41).

Fehler! Verweisquelle konnte nicht gefunden werden. zeigt insbesondere die steigende Anzahl an Unternehmen im Markt für darstellende Künste in Deutschland. So haben sich im Zeitraum von 2009 bis 2019 knapp 43% mehr Unternehmen im Markt der darstellenden Künste etablieren können als es noch im Jahr 2009 der Fall war. Die Auswirkungen der Weltfinanzkrise in den Jahren 2007 und 2008 und der Einschränkungen während der Corona Pandemie in den Jahren 2020 und 2021 stellten große Herausforderungen für den Markt dar. Dies unterstrich die Wichtigkeit einer nachhaltigen Wirtschaftsweise für die Betriebe.

Wirkungskette der Kundenbindung

Nach Bruhn 2005, S. 8–9, muss zur Erzielung von Kundenbindung eine Wirkungskette durchlaufen werden bis eine Kundenbindung entsteht. Die fünf Phasen dieser Wirkungskette werden im Folgenden dargestellt.

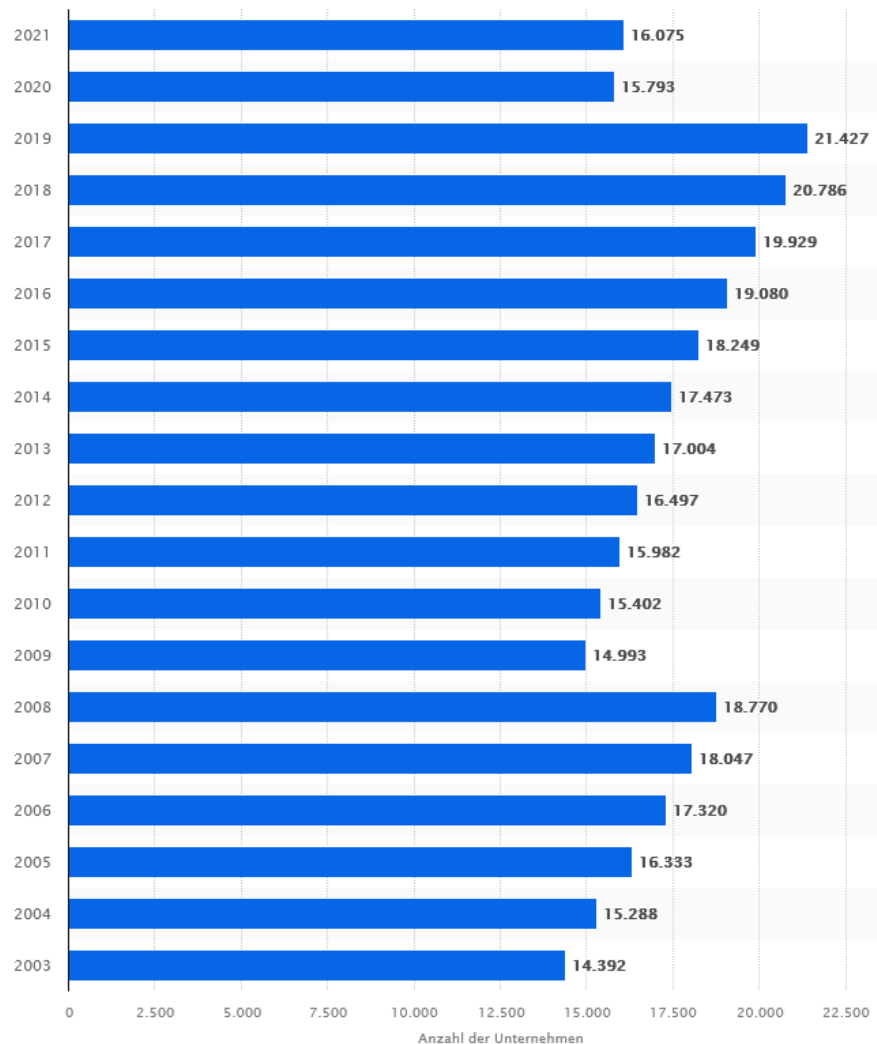


Abbildung 2: Anzahl der Unternehmen im Markt für darstellende Künste in Deutschland von 2003 bis 2021 (Statista 2023)

Der erste Schritt der Wirkungskette umfasst den Erstkontakt des Kunden mit dem Anbieter, entweder durch den Kauf eines Produkts oder die Nutzung einer Dienstleistung. Nach diesem Erstkontakt folgt Phase 2, in der der Kunde die Situation und Interaktion bewertet und sein persönliches Zufriedenheitsurteil bildet. Wenn diese Bewertung positiv ausfällt oder die Erwartungen des Kunden sogar übertroffen werden, kann in Phase 3 Kundenloyalität entstehen. Kundenloyalität umfasst ein grundlegendes Vertrauensverhältnis, eine allgemein positive Einstellung und die Akzeptanz der Leistungsfähigkeit des Anbieters seitens des Kunden. In dieser Situation zeigt der Kunde eine erhöhte Bereitschaft zum Wiederholungsgeschäft

und zur Empfehlung des Anbieters. Phase 4 markiert den Übergang zur Kundenbindung, wenn sich die positive Überzeugung des Kunden in tatsächlichem Wiederkaufverhalten, Cross-Buying oder Weiterempfehlungen an potenzielle Kunden widerspiegelt. In Phase 5 schließt sich die Wirkungskette mit einer Steigerung des ökonomischen Erfolgs aufgrund der erzielten Effekte. Der gesamte Ablauf der Wirkungskette wird von externen und internen Faktoren beeinflusst, die entweder positiv oder negativ auf den gewünschten Prozess wirken können.

Bei der Betrachtung dieser Wirkungskette fällt auf, dass Kundenbindung erst aus einem grundlegenden Vertrauensverhältnis des Kunden gegenüber dem Theater heraus entsteht. Insbesondere Daten zu der allgemeinen Einstellung des Kunden gegenüber des Theaters und seine Bereitschaft zur Weiterempfehlung könnte eine Modellierung der zukünftigen Kundenbindung stark verbessern.

Rolle des E-Mail-Marketings innerhalb der Wirkungskette der Kundenbindung

Es gibt viele Einflussfaktoren, die einen Übergang der Kunden in die nächste Phase begünstigen würden. Sowohl unternehmensexterne moderierende Faktoren wie die Heterogenität der Kundenerwartungen, das Variety-Seeking-Motive der Kunden oder die Anzahl an Alternativen auf dem Markt, als auch unternehmensinterne moderierende Faktoren, wie die Individualität des Theaterstücks, die Komplexität des Angebots eines Theaters oder persönliche Beziehungen der Mitarbeiter zu den Kunden haben Einfluss auf die verschiedenen Phasen. In dieser Arbeit interessieren wir uns dafür inwiefern die kundenbezogene Informationspolitik des Thalia Theaters in Form von E-Mail-Marketing Einfluss auf den Übergang von Phase 3 der Kundenloyalität zu Phase 4 der Kundenbindung haben kann.

3.2 Das Thalia Theater

Das Thalia Theater in Hamburg hat eine reiche Geschichte, die bis ins 19. Jahrhundert zurückreicht und nimmt mit den Spielstätten im Haupthaus und in der Gaußstraße unter den deutschsprachigen Sprechtheatern aufgrund seiner Qualität und Tradition eine herausgehobene Position ein. Das Thalia Theater betrachtet Theater als zeitgenössische Kunstform und engagiert sich für die Auseinandersetzung mit aktuellen gesellschaftlichen Themen. Sein Ziel ist es, ein breites Spektrum der dramatischen Literatur, einschließlich zeitgenössischer Werke, durch ein vielfältiges Angebot an Theateraufführungen, Workshops, Lesungen und ähnlichen Veranstaltungen auf hohem künstlerischem Niveau einem breiten Publikum zugänglich zu machen. Gleichzeitig strebt das Theater immer danach, neue Zuschauergruppen für das Theater zu gewinnen, insbesondere indem es Kinder und Jugendliche an die Welt des Theaters heranführt (Freie und Hansestadt Hamburg 2021, S. 35).

Geschichte des Thalia Theaters

Das frühere Gebäude des Thalia Theaters wurde 1843 gegenüber dem heutigen Bau errichtet. Es erhielt seinen Namen Thalia nach der Muse der komischen Dichtung und Unterhaltungskunst, die auch eine der drei Grazien ist. Im Jahr 1907 wurde das Thalia Theater in eine GmbH umgewandelt. 1912 wurde der neue Bau am heutigen Gerhart-Hauptmann-Platz mit 1.300 Sitzplätzen eröffnet. Im Jahr 1937 wurde das Thalia Theater verstaatlicht. Während des Zweiten Weltkrieges wurde es durch Bombenangriffe zerstört. Es wurde jedoch bereits im Jahr 1946 provisorisch wiedereröffnet. Im Dezember 1960 erfolgte die Eröffnung des restaurierten Hauses, dessen Innenausstattung als herausragendes Beispiel der Nachkriegsmoderne gilt. Im November 2000 wurde das Thalia in der Gaußstraße eröffnet, eine Studiobühne in Hamburg-Altona mit 200 Plätzen. Dort finden hauptsächlich Erstaufführungen junger Autoren und experimentelles Theater ihren Platz.

Heute ist das Thalia Theater ein staatliches Theater mit einem festen Ensemble. Das Große Haus bietet etwa 1.000 Plätze. Das Repertoire umfasst rund 20 Produktionen, die entweder täglich wechseln oder in Blöcken aufgeführt werden. Jährlich gibt es etwa neun Premieren im Großen Haus und sechs Premieren im Thalia in der Gaußstraße (hamburg.de 2023).

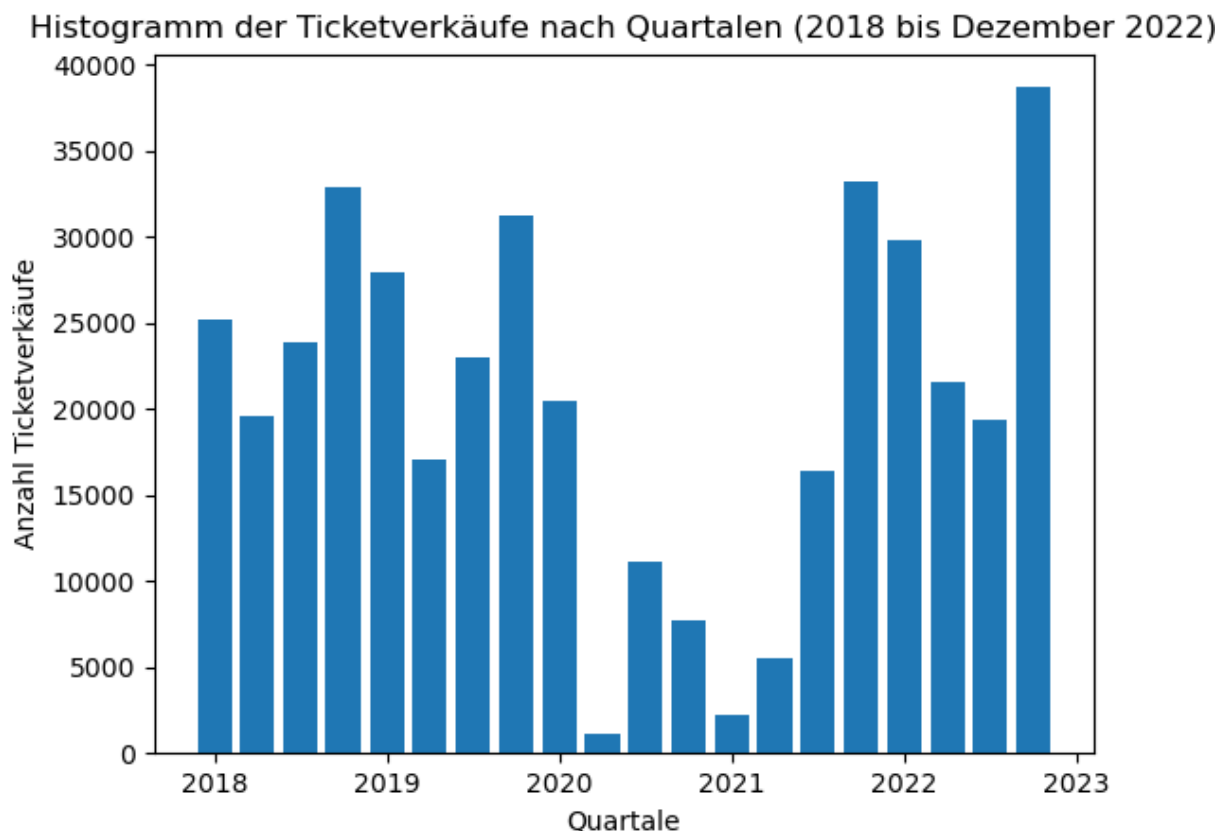


Abbildung 3: Ticketverkäufe des Thalia Theaters in den Jahren 2018 bis 2022 nach Quartalen

Wirtschaftlichkeit

Der Kostendeckungsgrad des Thalia Theaters betrug im Jahr 2018 22,5% und stieg im Jahr 2019 leicht auf 22,8% (Freie und Hansestadt Hamburg 2021, S. 45). Dies zeigt deutlich, dass das Thalia Theater für den Erhalt des Betriebs stark auf die Förderungen der Stadt Hamburg angewiesen ist. Die Behörde für Kultur und Medien der Stadt Hamburg formuliert wiederum klare Ziele an das Thalia Theater, um eine öffentliche Förderung zu rechtfertigen. Diese Ziele werden an drei Kennzahlen gemessen und umfassen sowohl die Gesamtanzahl der Besucher/-innen und Anzahl der juvenilen/ jugendlichen Besucher, als auch den Kostendeckungsgrad. So wurde ein Kostendeckungsgrad von 23,0% als Ziel für das Jahr 2022 ausgegeben. Für die Jahre 2023-2026 wurde jeweils ein Kostendeckungsgrad von 21,0% als Zielwert festgelegt. Im Jahr 2021 wurde unter anderem aufgrund von Beschränkungen während der Corona Pandemie eine Kostendeckung von 7,20% erreicht (Freie und Hansestadt Hamburg 2022, S. 47).

Saisonaler Effekt

Abbildung 4 zeigt den für den Theatermarkt typischen saisonalen Effekt. Dieser ist dadurch gekennzeichnet, dass ein Großteil des Umsatzes in den Wintermonaten erzielt wird.

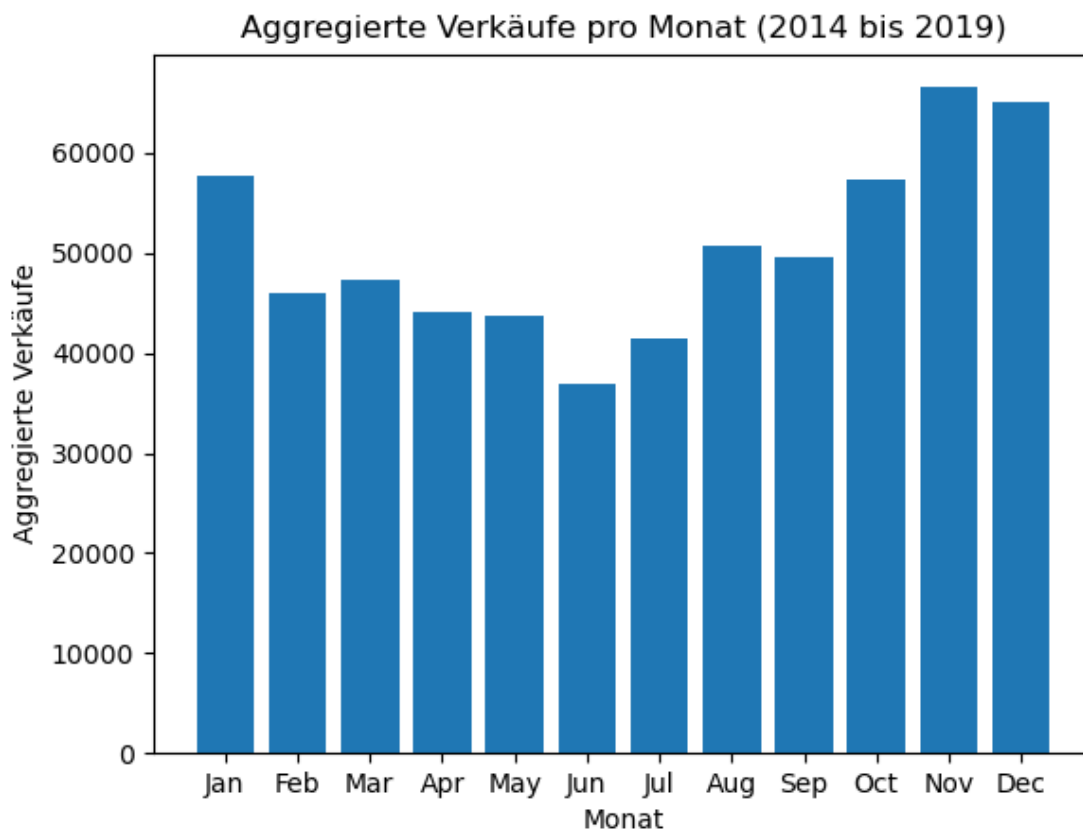


Abbildung 4: Aggregierte Verkäufe pro Monat (2014 bis 2019)

3.3 E-Mail-Marketing des Thalia Theaters

Das folgende Unterkapitel untersucht das E-Mail-Marketing des Thalia Theaters im Zeitraum von Oktober 2021 bis Dezember 2022. Dabei liegt der Fokus auf zwei zentralen Aspekten: Der Häufigkeit des E-Mail-Versands und der Gestaltung der E-Mails. Durch die Analyse dieser Themen werden Einblicke in die E-Mail-Kommunikation des Theaters gewonnen und deren Auswirkungen auf die Empfänger untersucht.

Häufigkeit des E-Mail-Versands

Im Zeitraum von Oktober 2021 bis Dezember 2022 wurden insgesamt 30 Kampagnen vom Thalia Theater durchgeführt, um das E-Mail-Marketing zu gestalten und die Zielgruppe effektiv anzusprechen. Abbildung 5 visualisiert die Größe und das Muster der Kampagnenaktivitäten des Thalia Theaters. Im Durchschnitt wurden pro Monat etwa zwei Kampagnen geschaltet, wobei oftmals auf einen zeitlichen Abstand von zwei Wochen zwischen den Kampagnen geachtet wurde. Auffallend sind die beiden Kampagnen im März 2022 als eine zusätzliche Kampagne zur „Solidarität der Ukraine“ geschaltet wurde.

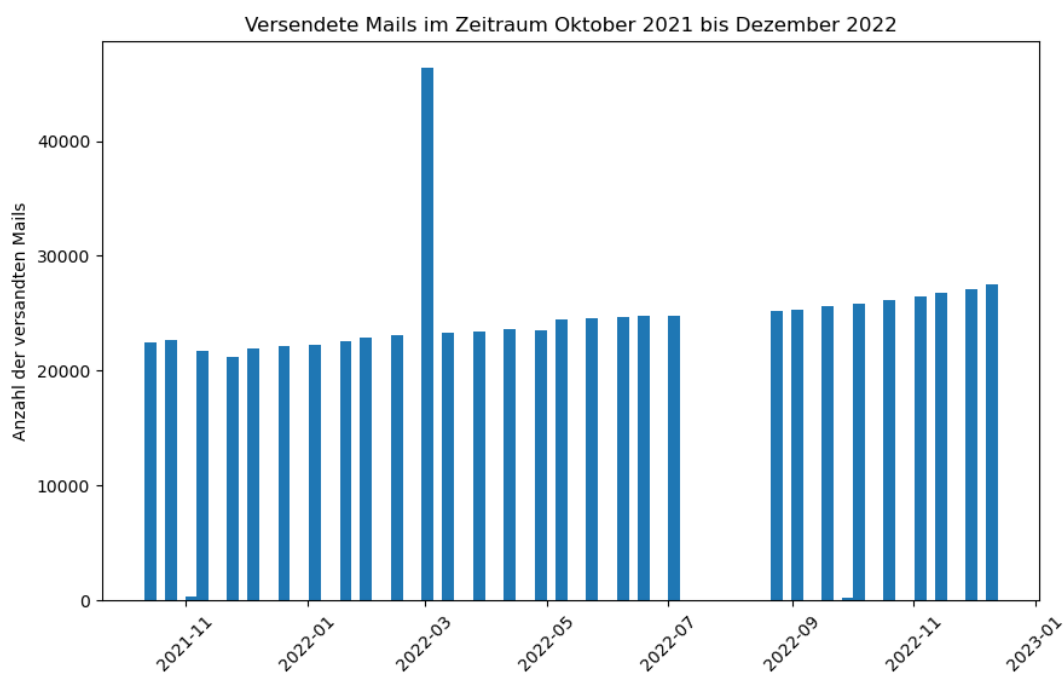


Abbildung 5: Versendete E-Mails im Zeitraum Oktober 2021 bis Dezember 2022

Gestaltung der E-Mail-Kampagnen

Die E-Mail-Kampagnen des Thalia Theaters folgen in der Regel folgendem Muster:

1. Ankündigung bevorstehender Veranstaltungen im Thalia Theater: Informationen über kommende Veranstaltungen und Aufführungen im Theater.

2. Hervorhebung der Künstler und Stücke: Detaillierte Vorstellung der Künstler und der Inhalte der Stücke, um das Interesse der Leser zu wecken.
3. Verwendung von Zitaten und positiven Bewertungen: Zitate positiver Bewertungen oder Kommentare, um das Publikum von der Qualität der Aufführungen zu überzeugen und Vorfreude zu erzeugen.
4. Sonderaktionen und Rabatte: Motivation des Publikums zum Ticketkauf durch Rabatte und Aktionen.
5. Einbindung des Publikums: Ermutigung des Publikums zur Teilnahme an Umfragen oder zur Interaktion mit dem Theater, um Feedback zu erhalten und die Theatererfahrung zu verbessern.
6. Ankündigung bevorstehender Veranstaltungen im Thalia Gauß: Links und Informationen zum Kauf von Tickets.
7. Spielplan für die nächsten 14 Tage im Thalia Theater: Links und Informationen zum Kauf von Tickets.

3.4 Öffnungserkennung

Die Erkennung einer E-Mail-Öffnung kann technisch realisiert werden, indem ein Tracking-Pixel, ein kleines, unsichtbares Bild, in die E-Mail eingebettet wird. Aufgrund des verzögerten Nachladens aller Bilder, einschließlich des Tracking-Pixels, kann die E-Mail-Marketing-Software eine Öffnung feststellen, sobald das Mail-Programm die Anforderung für das Laden des Tracking-Pixels sendet. Dieser Prozess bleibt für den Empfänger in der Regel unbemerkt, da seine Erkennung nicht offensichtlich ist (vgl. Kornfeld 2023).

Wie zuverlässig ist die Öffnungs-Erkennung?

Laut Kornfeld 2023 ist die exakte Bestimmung der Öffnungsrate nicht zu 100% präzise, da verschiedene Faktoren zu einer Reihe möglicher Verzerrungen führen können, die sowohl positive als auch negative Auswirkungen haben können. Im Folgenden werden mehrere mögliche Faktoren für Verzerrungen, die für unsere Analysen relevant sind, aufgeführt:

Vorschau-Fenster: Wenn eine E-Mail nur im Vorschau-Fenster angezeigt wird, kann das Tracking-Pixel geladen und als Öffnung gewertet werden, obwohl der Empfänger die E-Mail tatsächlich nicht gelesen hat.

Sichtbarer Bereich: In einigen E-Mail-Programmen wird das Tracking-Pixel nur dann geladen, wenn es sich im sichtbaren Bereich befindet. Dadurch hängt die Erkennung einer Öffnung beispielsweise von der Position des Tracking-Pixels in der E-Mail ab.

Löschungs-Klick: Wenn eine E-Mail lediglich angeklickt wird, um sie sofort danach zu löschen, kann sie möglicherweise (kurzzeitig) im Vorschau-Fenster angezeigt werden, und somit wird eine Öffnung festgestellt, die in der Praxis tatsächlich nicht stattgefunden hat.

Apple Mail Privacy Protection seit 20. September 2021: Die Funktion "Mail Privacy Protection" in der Mail-App von Apple ist standardmäßig deaktiviert. Per Opt-in verhindert sie die Verwendung von Pixeln zur Erfassung von Informationen über den Empfänger. Dies umfasst auch die Erkennung, ob eine E-Mail geöffnet wurde oder nicht (vgl. Stewart 2021).

Gmail Privacy Protection seit 2013: Vor dem Versand einer E-Mail an einen Abonnenten lädt der Gmail-Server die enthaltenen Bilder. Dieser Prozess kann dazu führen, dass E-Mail-Marketingplattformen falsch-positive Ergebnisse übermitteln (vgl. Stewart 2021).

3.5 Nutzen der E-Marketing Suite der Firma Wilken

Am Beispiel des Theater Basel lässt sich gut veranschaulichen, welche Rolle die E-Marketing Suite der Firma Wilken bei der Gestaltung von E-Mail-Kampagnen bei Kulturunternehmen einnimmt. Die E-Marketing Suite bietet Schnittstellen zu anderen Anbietern im CRM-Markt an, wie in **Fehler! Verweisquelle konnte nicht gefunden werden.** dargestellt. So erfasst das Ticketsystem der Firma Secutix während des Ticketing-Prozesses die Transaktionsdaten und übermittelt diese an die Firma kulturplaner. Kulturplaner wiederum verarbeitet die Daten mit Methoden der Business Intelligence und transportiert die Daten weiter zur E-Marketing Suite. Die E-Marketing Suite speichert dann alle Kunden- und Transaktionsdaten. Kundendaten werden in Echtzeit übertragen und Transaktionsdaten aufgrund der großen Datenmenge einmal täglich.

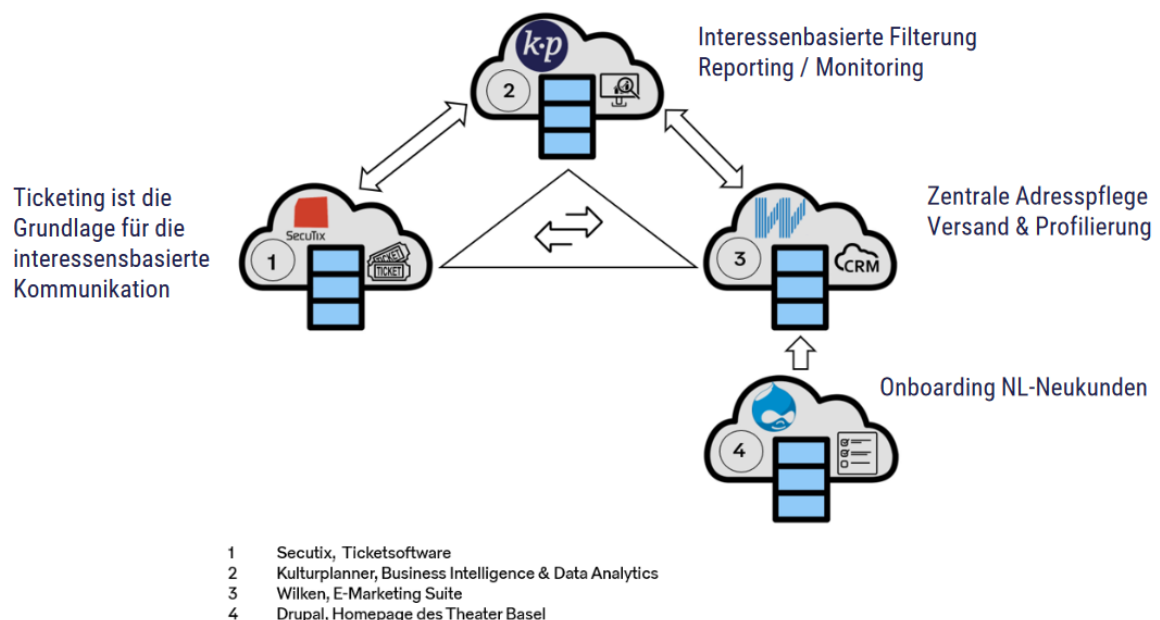


Abbildung 6: Einbindung der E-Marketing Suite in die CRM-Landschaft des Theater Basel (Franziska Pinkert et al. 2022)

Die E-Marketing Suite fungiert als zentrale und nachhaltige Adressdatenbank innerhalb der CRM-Landschaft. Unter anderem die Pflege der Adressen finden über die E-Marketing Suite statt, wobei beliebig viele User an den Daten arbeiten können. Des Weiteren werden Newsletter-Neuanmeldungen direkt von der Homepage des Theaters an die E-Marketing Suite übermittelt. Wilken bietet für diese Neuanmeldungen Formulare an, die das Theater auf seiner Homepage nutzen kann. Da der Versand der E-Mail-Kampagnen ebenfalls über die E-Marketing Suite verläuft, enthalten die Kundendaten in der E-Marketing Suite Informationen zu den Öffnungen und Klicks sowie Informationen, wann und ob eine E-Mail erhalten wurde.

4 Das Hidden Markov Model

Durch die Modellierung von verborgenen Zuständen und beobachtbaren Ereignissen liefert das HMM wertvolle Einblicke, um effektive Marketingstrategien zu entwickeln, gezielte Werbung zu ermöglichen und Kundenbeziehungen zu verbessern. In diesem Kapitel wird zunächst eine Intuition für das verwendete Modell mit Hilfe eines Münzwurfbeispiels vermittelt. Anschließend werden die charakterisierenden Eigenschaften eines Hidden Markov Models dargestellt. Abschließend wird beschrieben wie das Hidden Markov Model für unsere Modellierung verwendet wird.

4.1 Intuition

Dadurch, dass das Modell anschließend Anwendung in der Praxis finden soll, wird in diesem Unterkapitel ein großer Wert darauf gelegt, einen intuitiven Zugang zu dem Modell zu vermitteln. Es wurde ein Münzwurf-Beispiel zur Veranschaulichung gewählt.

Ein Hidden Markov Model (HMM) ist ein zweifacher stochastischer Prozess mit einem zugrunde liegenden stochastischen Prozess, der „verborgen“ ist, weil er nur durch eine weitere Reihe von stochastischen Prozessen beobachtet werden kann, die die Abfolge der beobachteten Symbole erzeugen. Wir veranschaulichen HMMs anhand des folgenden Beispiels mit Münzwürfen (vgl. Rabiner und Juang 1986, S. 5–7).

Um das Konzept des HMM zu verstehen, betrachten wir das folgende vereinfachte Beispiel: Man befindet sich in einem Raum mit einer Barriere, beispielsweise einem Vorhang, durch den man nicht sehen kann, was auf der anderen Seite passiert. Auf der anderen Seite der Barriere befindet sich eine andere Person, die ein Münzwurf-Experiment (oder mehrere Münzwurf-Experimente) durchführt. Die andere Person wird nichts über ihre genauen Handlungen verraten; sie wird nur das Ergebnis jedes Münzwurfs mitteilen. Es werden also eine Reihe verborgener Münzwurfexperimente durchgeführt und nur die Ergebnisse der Münzwürfe beobachtet, beispielsweise Beobachtung = Kopf, Kopf, Zahl, Kopf, Zahl, Zahl, Kopf, Kopf, ..., Zahl.

Bei dem oben beschriebenen Experiment besteht die Herausforderung darin, ein HMM zu erstellen, das die beobachtete Abfolge von "Kopf" und "Zahl" erklärt. Ein mögliches Modell ist in Abbildung 7a) dargestellt. Wir nennen dies das "1-fair coin model". Das Modell besteht aus zwei Zuständen, wobei jedem Zustand eindeutig entweder "Kopf" (Zustand 1) oder "Zahl" (Zustand 2) zugeordnet ist. Daher ist dieses Modell nicht verborgen, da die Beobachtungssequenz den Zustand eindeutig bestimmt. Das Modell repräsentiert eine "faire Münze", da die Wahrscheinlichkeit, nach einem "Kopf" (oder "Zahl") einen "Kopf" (oder "Zahl") zu erzeugen, 0,5 beträgt. Es gibt also keine Verzerrung in Bezug auf die aktuelle Beobachtung. Dies ist ein vereinfachtes Beispiel und zeigt, wie unabhängige Versuche wie das Werfen einer fairen Münze als eine Abfolge von aufeinanderfolgenden Ereignissen interpretiert werden können.

Natürlich sollte dieses Modell, wenn die Person hinter der Barriere tatsächlich eine faire Münze wirft, die Ergebnisse sehr gut erklären.

Ein weiteres mögliches HMM zur Erklärung der beobachteten Abfolge von Münzwurfresultaten wird in Abbildung 7b gezeigt. Wir nennen dieses Modell das "2-fair coins model". Es gibt wiederum 2 Zustände im Modell, aber kein Zustand ist eindeutig mit entweder "Kopf" oder "Zahl" assoziiert. Die Wahrscheinlichkeit für "Kopf" (oder "Zahl") in jedem Zustand beträgt 0,5. Auch die Wahrscheinlichkeit, den Zustand zu wechseln (oder im aktuellen Zustand zu bleiben), beträgt 0,5. Somit können wir in diesem Fall jedem Zustand eine faire (unverzerrte) Münze zuordnen. Obwohl die Wahrscheinlichkeiten für das Verbleiben in oder das Verlassen eines der beiden Zustände jeweils 0,5 sind, sollte der Leser nach etwas Überlegung erkennen, dass die Statistiken der beobachtbaren Ausgangssequenzen des "2-fair coins model" unabhängig von den Zustandsübergängen sind. Der Unterschied zum „1-fair coin model“ ist, dass dieses Modell verborgen ist (d. h. wir können nicht genau wissen, welche faire Münze (Zustand) zu den beobachteten "Kopf" oder "Zahl" bei jeder Beobachtung geführt hat), aber im Wesentlichen nicht von dem "1-fair coin model" in Abbildung 7a zu unterscheiden ist.

In den Abbildungen 1c und 1d werden zwei weitere mögliche HMMs gezeigt, die die beobachtete Abfolge von "Kopf" und "Zahl" erklären können. Das Modell in Abbildung 1c, das wir das "2-biased coins model" nennen, hat zwei Zustände (entsprechend zwei verschiedenen Münzen). Im Zustand 1 ist die Münze stark auf "Kopf" verzerrt, im Zustand 2 ist die Münze stark auf "Zahl" verzerrt. Die Übergangswahrscheinlichkeiten zwischen den Zuständen sind alle gleich 0,5. Dieses "2-verzerrte Münzen"-Modell ist ein verborgenes Markov-Modell, das sich von den zuvor diskutierten Modellen unterscheidet. Interessanterweise sollte der Leser sich selbst davon überzeugen können, dass die Statistiken (z. B. durchschnittliche Anzahl von "Kopf" oder "Zahl") der Beobachtungssequenzen des HMMs in Abbildung 1c auf lange Sicht dieselben sind wie die der Modelle in den Abbildungen 7a und 7b. Dieses Modell ist sehr geeignet, wenn sich hinter der Barriere folgendes abspielt: Die Person hat drei Münzen, eine faire und zwei verzerrte Münzen entsprechend der Beschreibung in Abbildung 7c. Die beiden verzerrten Münzen sind jeweils mit den beiden Seiten der fairen Münze assoziiert. Um das Ergebnis jedes Münzwurfs zu erhalten, wirft die Person hinter der Barriere zuerst die faire Münze, um zu entscheiden, welche verzerrte Münze verwendet werden soll, und wirft dann die ausgewählte verzerrte Münze, um das Ergebnis zu erhalten. Mit diesem Modell sind wir in der Lage, die oben genannten subtilen Änderungen der statistischen Merkmale (d.h. den Wechsel der verzerrten Münzen) zu betrachten und zu erklären.

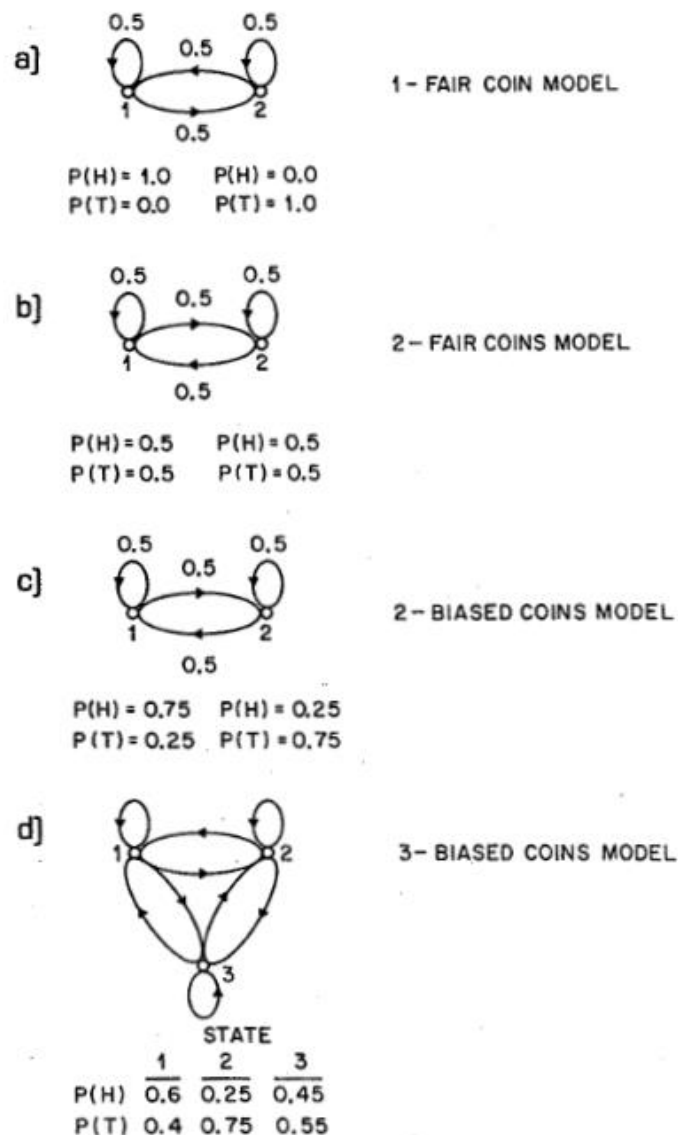


Abbildung 7: Modellierung von verborgenen Münzwurf-Experimenten (Rabiner und Juang 1986, S. 6)

Das Modell in Abbildung 7d, das wir das "3-biased coins model" nennen, hat drei Zustände (entsprechend drei verschiedenen Münzen). Im Zustand 1 ist die Münze leicht auf "Kopf" verzerrt, im Zustand 2 ist die Münze stark auf "Zahl" verzerrt und im Zustand 3 ist die Münze leicht auf "Zahl" verzerrt. In Abbildung 7d sind die Übergangswahrscheinlichkeiten der Zustände nicht spezifiziert. Es ist klar, dass das Verhalten der beobachteten Sequenzen, die von einem solchen Modell erzeugt werden, stark von diesen Übergangswahrscheinlichkeiten abhängt. (Um sich davon zu überzeugen, sollte man zwei Extreme betrachten, nämlich wenn die Wahrscheinlichkeit, im Zustand 3 zu bleiben, groß ($>0,95$) oder klein ($<0,05$) ist. Aufgrund der starken Verzerrung der Münze, die dem Zustand 3 zugeordnet ist, ergeben sich aus diesen beiden Extremen sehr unterschiedliche Sequenzstatistiken.) Wie beim "3-biased coins model" kann

hinter der Barriere ein reales Szenario existieren, das diesem Modell entspricht. Der Leser sollte keine Schwierigkeiten haben, ein solches Szenario selbst zu entwickeln.

Aus dieser Diskussion über die Modellierung der Ausgänge des Münzwurfs mit HMMs lassen sich mehrere wichtige Punkte ableiten. Zunächst einmal ist festzustellen, dass eine der schwierigsten Aufgaben des Modellierungsverfahrens darin besteht, die Größe des Modells (die Anzahl der Zustände) festzulegen. Ohne vorherige Informationen ist diese Entscheidung oft schwierig zu treffen und kann mehrere Versuche erfordern, bevor die passendste Modellgröße gefunden wird. Obwohl wir in der obigen Darstellung bei einem 3-Münzen-Modell gestoppt haben, könnte selbst dies zu klein sein. Wie entscheiden wir, wie viele Münzen (Zustände) tatsächlich im Modell benötigt werden? Die Antwort auf diese Frage hängt mit einer noch größeren Frage zusammen, nämlich wie wir die Modellparameter (Übergangswahrscheinlichkeiten, Wahrscheinlichkeiten für "Kopf" und "Zahl" in jedem Zustand) wählen, um das Modell zu optimieren, sodass es die beobachtete Ergebnissequenz am besten erklärt. In dem Abschnitt über Lösungen für die drei HMM-Probleme werden wir versuchen, diese Frage zu beantworten, da dies der Schlüssel für den erfolgreichen Einsatz von HMMs bei realen Problemen ist.

Ein letzter Punkt betrifft die Größe der Beobachtungssequenz. Wenn wir auf eine kleine, endliche Beobachtungssequenz beschränkt sind, können wir möglicherweise die optimalen Modellparameter nicht zuverlässig schätzen. (Denken Sie an den Fall, dass tatsächlich 10 Münzen verwendet werden, aber nur 50-100 Beobachtungen gegeben sind.) Je nach der Menge an Trainingsdaten für das Modell können bestimmte HMMs möglicherweise statistisch nicht zuverlässig unterschieden werden.

4.2 Elemente eines Hidden Markov Models

Wir erklären nun die Elemente und den Mechanismus des HMMs, das wir in dieser Arbeit behandeln (vgl. Rabiner und Juang 1986, S. 7):

1. Es gibt eine endliche Anzahl von Zuständen in dem Modell, wir bezeichnen diese als $S = 1, \dots, NS$ (für „state“, dt. „Zustand“). Jeder Zustand spiegelt ein anderes Niveau der Kundenbindung wider.
2. Zu jedem Zeitpunkt t tritt ein neuer Zustand basierend auf einer Übergangswahrscheinlichkeitsverteilung ein, die ausschließlich vom vorherigen Zustand abhängt (Eigenschaft des Markow-Prozesses). (Es sei darauf hingewiesen, dass der Übergang so sein kann, dass der Prozess im vorherigen Zustand bleibt.)
3. Nach jedem Übergang wird eine Interaktion des Kunden mit dem Unternehmen gemäß einer Wahrscheinlichkeitsverteilung erzeugt, die ausschließlich vom aktuellen Zustand abhängt. Diese Wahrscheinlichkeitsverteilung bleibt unabhängig davon, wann und wie der Kunde den Zustand betreten hat, für den Zustand konstant. Es gibt also N solcher

Beobachtungswahrscheinlichkeitsverteilungen, die natürlich stochastische Prozesse repräsentieren

5 Beschreibung des verwendeten Modells

Im Folgenden wird das verwendete Modell beschrieben. Gegenüber dem Modell von Zhang et al. (2017) wird in dem Modell eine Poisson-Verteilung statt einer Binomialverteilung für die Modellierung der Öffnungszahlen verwendet. Dies hat den Hintergrund, dass Kunden auch mehr E-Mails öffnen können als sie in diesem Monat erhalten haben, da sie E-Mails aus vergangenen Monaten ebenfalls ansehen können. Des Weiteren wird die Endogenitätskomponente bei der Definition der Parameter dargestellt, um diese Arbeit als mögliche Referenz in zukünftigen Anwendungen verwendet werden kann. Bei der Berechnung in dieser Arbeit, werden die Endogenitätskomponenten nicht verwendet, um die Komplexität der Berechnungen zu reduzieren.

Sei O_{it} die Anzahl der E-Mails, die Kunde i zur Zeit t öffnet, und Y_{it} die Anzahl der Einkäufe, die Kunde i zur Zeit t tätigt. Für Kunden i modellieren wir die Abfolge der Beobachtungen $[(Y_{i1} = y_{i1}, O_{i1} = o_{i1}), \dots, (Y_{it} = y_{it}, O_{it} = o_{it})]$ unter Verwendung eines HMM, das durch (1) die Anfangszustandsverteilung (ψ_i), (2) eine Sequenz von Übergangswahrscheinlichkeiten (Q_{it}) und (3) einen Vektor von Wahrscheinlichkeiten, die die latenten Zustände mit den beobachteten Kauf- und E-Mail-Öffnungsverhalten in Beziehung setzen (H_{it}), charakterisiert wird.

5.1 Startverteilung

Zu einem beliebigen Monat t sei s der Grad der Beziehung des Kunden i . Sei ψ_{is} die Wahrscheinlichkeit, dass Kunde i zu Beginn in Beziehungszustand s ist, wobei $\psi_{is} \geq 0$ und $\sum_{s=1}^{NS} \psi_{is} = 1$, wobei NS die Anzahl der latenten Markov-Zustände ist. In dieser Arbeit gehen wir davon aus, dass alle Kunden im ersten Monat mit dem niedrigsten Beziehungszustand beginnen. Daher ist $\psi_{is} = (\psi_{i1}, \dots, \psi_{i,NS}) = (1, 0, \dots, 0)$.

5.2 Übergangsmatrix

In unserem vorgeschlagenen HMM-Framework erlauben wir Kunden, in jeden Beziehungszustand $s = 1, 2, \dots, NS$ zu wechseln. Wir verwenden eine multinominale Logit-Spezifikation, um diesen Übergangsprozess zu formulieren. Wir definieren die Übergangsmatrix wie folgt:

$$Q_{i,t-1 \rightarrow t} = \begin{array}{c|cccc} & \text{State} & 1 & 2 & \dots & NS \\ \hline 1 & q_{it,1 \rightarrow 1} & q_{it,1 \rightarrow 2} & \dots & q_{it,1 \rightarrow NS} \\ 2 & q_{it,2 \rightarrow 1} & q_{it,2 \rightarrow 2} & \dots & q_{it,2 \rightarrow NS} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ NS & q_{it,NS \rightarrow 1} & q_{it,NS \rightarrow 2} & \dots & q_{it,NS \rightarrow NS} \end{array},$$

Hierbei ist $q_{it,s \rightarrow s'}$ die Wahrscheinlichkeit, dass Kunde i vom Zustand s zur Zeit $t-1$ in den Zustand s' zur Zeit t wechselt, wobei $0 \leq q_{it,s \rightarrow s'} \leq 1$ für alle s, s' gilt und $\sum_{s'=1}^{NS} q_{it,s \rightarrow s'} = 1$ erfüllt ist.

Wir spezifizieren den indirekten Übergangsnutzen des Kunden i für den Wechsel vom Beziehungszustand s im Zeitraum $t-1$ in den Zustand s' zur Zeit t ($U_{it,s \rightarrow s'}$) wie folgt:

$$u_{it,s \rightarrow s'} = v_{it,s \rightarrow s'} + e_{it,s \rightarrow s'} \quad (1)$$

Hierbei bezeichnen $e_{it,s \rightarrow s'}$ und $V_{it,s \rightarrow s'}$ die stochastischen und deterministischen Komponenten des indirekten Nutzens. Wir nehmen an, dass $e_{it,s \rightarrow s'}$ für $i = 1, \dots, N$; $t = 1, \dots, T$; $s, s' = 1, \dots, NS$ i.i.d. Gumbel-verteilt mit Lageparameter 0 und Skalenparameter 1 sind. Wir operationalisieren die deterministische Komponente $v_{it,s \rightarrow s'}$ wie folgt:

$$v_{it,s \rightarrow s'} = \alpha_{s \rightarrow s'} + \gamma_{s \rightarrow s'} X_{i,t-1} \quad (2)$$

wobei $\alpha_{s \rightarrow s'}$ der intrinsische Nutzen ist, der mit der Transition von Zustand s zu s' zum Zeitpunkt t einhergeht. $X_{i,t-1}$ enthält folgende Variablen: $I[O_{i,t-1} > 0]$, $I[Y_{i,t-1} > 0]$, $EM_{i,t-1}$, $EM_{i,t-1}^2$, wobei $I[A]$ die Indikatorfunktion ist, die den Wert 1 annimmt, wenn das Ereignis A eintritt, und sonst den Wert 0 annimmt. $O_{i,t-1}$ ist die verzögerte Anzahl geöffneter E-Mails, $Y_{i,t-1}$ ist die verzögerte Anzahl von Käufen und $EM_{i,t-1}$ ist die Anzahl der vom Unternehmen zum Zeitpunkt $t-1$ gesendeten E-Mails. $\gamma_{s \rightarrow s'} = [\gamma_{1,s \rightarrow s'}, \gamma_{2,s \rightarrow s'}, \gamma_{3,s \rightarrow s'}, \gamma_{4,s \rightarrow s'}]$ bezeichnet den Vektor der entsprechenden Antwortkoeffizienten. Wir normalisieren den deterministischen Nutzen für Kunden i zum Zeitpunkt t , um zur niedrigsten Stufe ($V_{it,s \rightarrow 1}$) den Wert 0 für Identifikationszwecke zu erhalten (d. h., $U_{it,s \rightarrow 1} = e_{it,s \rightarrow 1}$, wobei $e_{it,s \rightarrow 1}$ für $i = 1, \dots, N$; $t = 1, \dots, T$; $s = 1, \dots, NS$ i.i.d. Gumbel-verteilt mit Lageparameter 0 und Skalenparameter 1 sind). Daher wird die Übergangswahrscheinlichkeit für Kunden i , der von Zustand s zu s' zum Zeitpunkt t übergeht ($q_{it,s \rightarrow s'}$), zur folgenden bekannten Multinomial-Logit-Share-Funktion:

$$q_{it,s \rightarrow s'} = \frac{\exp(v_{it,s \rightarrow s'})}{1 + \sum_{k=2}^{NS} \exp(v_{it,s \rightarrow k})} \quad (3)$$

Als nächstes diskutieren wir unser bedingtes E-Mail-Öffnungszählmodell (CEOM) und unser Kaufzählmodell (CPM).

5.3 Bedingte E-Mail-Öffnungsanzahl (CEOM)

Wir nehmen an, dass die Anzahl der E-Mails, die Kunde i (im Zustand s) zum Zeitpunkt t öffnet, einer Poissonverteilung mit Parameter λ_{it} folgt. P sei definiert als

$$\rho = E[E - \text{Mail} - \text{Öffnungen}] \quad (4)$$

Wir modellieren $\lambda_{it|s}$ als Funktion des vergangenen E-Mail-Öffnungsverhaltens der Kunden wie folgt:

$$\lambda_{it|s} = \exp [\rho + \alpha_{io|s} + \beta_{io} \ln(LO_{it})] \quad (5)$$

wobei $\alpha_{io|s}$ und β_{io} den intrinsischen Nutzen des Öffnens einer E-Mail und den Einfluss des zeitlichen Abstands zwischen Käufen (im Zustand s) erfassen. Wir erfassen die Dauerabhängigkeit durch die Variable der Zeit seit der letzten geöffneten E-Mail (d. h. LO_{it}). Wir verwenden den natürlichen Logarithmus von LO_{it} , um den abnehmenden Effekt zu erfassen. Zur Identifikation legen wir die folgenden Beschränkungen fest: (1) $\alpha_{io|s+1} = \alpha_{io|s} + \exp(\nabla \alpha_{io|s+1})$, wobei $\nabla \alpha_{io|s+1}$ ein Parameter ist, der aus den Daten geschätzt wird, und (2) β_{io} ist zustandsinvariant. Diese beiden Einschränkungen garantieren, dass Kunden in einem höheren Beziehungszustand bei gleichen Bedingungen eine höhere Wahrscheinlichkeit haben, E-Mails zu öffnen als Kunden in einem niedrigeren Zustand. Zusätzlich erlauben wir sowohl $\alpha_{io|s}$ als auch β_{io} kundenspezifisch zu sein, um die unbeobachtete Kundenheterogenität zu kontrollieren. Wir nehmen an, dass $\alpha_{io|s}$ und β_{io} mit den folgenden Normalverteilungen verteilt sind.

$$\begin{aligned} \alpha_{io|s} &= \alpha_{o|s} + \nabla \alpha_{io} \text{ und} \\ \beta_{io} &= \beta_o + \nabla \beta_{io}, \end{aligned} \quad (6)$$

wobei $\nabla \alpha_{io} \sim N(0, \sigma_{\alpha_o}^2)$ und $\nabla \beta_{io} \sim N(0, \sigma_{\beta_o}^2)$

5.4 Modell der Kaufanzahl (CPM)

Unter der Bedingung, dass sich der Kunde i zum Zeitpunkt t im Beziehungszustand s befindet, nehmen wir an, dass die Anzahl der Einkäufe, die der Kunde tätigt, einer ZINBD mit Parametern $\varphi_{it|s}$, $\lambda_{it|s}$ und r folgt. Für jede Beobachtung y_{it} nimmt ZINBD an, dass es zwei datengenerierende Prozesse gibt (basierend darauf, ob das Ergebnis gleich Null oder größer als Null ist), die wie folgt definiert sind:

$$P(Y_{it} = y_{it}|s) = \begin{cases} \varphi_{it|s} + (1 - \varphi_{it|s}) \left(1 + \frac{\lambda_{it|s}}{r}\right)^{-r} & \text{falls } y_{it} = 0 \\ (1 - \varphi_{it|s}) \frac{\Gamma(y_{it} + r)}{y_{it}! \Gamma(r)} \left(1 + \frac{\lambda_{it|s}}{r}\right)^{-r} \left(1 + \frac{r}{\lambda_{it|s}}\right)^{-y_{it}} & \text{falls } y_{it} > 0 \end{cases} \quad (7)$$

wobei r ein Streuparameter ist, für den angenommen wird, nicht von Kovarianten abhängig zu sein. Der bedingte Mittelwert und die Varianz des ZINBD sind gegeben durch $E(y_{it|s}) = (1 - \varphi_{it|s})\lambda_{it|s}$ und $\text{Var}(y_{it|s}) = (1 - \varphi_{it|s})\lambda_{it|s}(1 + \varphi_{it|s}\lambda_{it|s} + \lambda_{it|s}r)$. $\varphi_{it|s}$ und $\lambda_{it|s}$ erfassen die bedingte Null-Infations-Wahrscheinlichkeit und den bedingten erwarteten Kauf des Kunden i (im Zustand s) zum Zeitpunkt t .

Um den Überschuss an Nicht-Käufen zu berücksichtigen, modellieren wir die bedingte Null-Infations-Komponente der ZINBD ($\varphi_{it|s}$) als:

$$\varphi_{it|s} = \frac{1}{1 + \exp[\delta_{0i|s} + \delta_{1i|s} \ln(LY_{it})]} \quad (8)$$

wobei $\delta_{0i|s}$ und $\delta_{1i|s}$ den intrinsischen Nutzen des Kaufs und den Effekt der Dauerabhängigkeit im Zustand s wiedergeben. Wir erfassen die Dauerabhängigkeit durch die Variable Zeit seit dem letzten Kauf (LY_{it}). Wir verwenden den natürlichen Logarithmus von LY_{it} , um den abnehmenden Effekt zu erfassen. Um die unerfasste Heterogenität zu berücksichtigen, erlauben wir, dass $\delta_{ki|s}$ (für $k=0,1$) kundenspezifisch sind. Wir nehmen an, dass $\delta_{ki|s}$ (für $k=0,1$) über Kunden wie folgt normal verteilt sind:

$$\delta_{ki|s} = \delta_{k|s} + \nabla\delta_{ki}, k = 0, 1 \quad (9)$$

wobei $\nabla\delta_{ki} \sim N(0, \sigma_{ki}^2)$ für $k = 0, 1$.

Wir modellieren die zweite Komponente des bedingten Kaufmodells (d. h. bedingter erwarteter Anzahl an Käufen [$\lambda_{it|s}$]) als eine Funktion der Anzahl der E-Mails, die von der Firma gesendet wurden (EM_{it}), gegeben durch

$$\lambda_{it|s} = \exp(\alpha_{ip|s} + \beta_{1,ip|s}EM_{it} + \beta_{2,ip|s}EM_{it}^2) \quad (10)$$

Bedingt auf den Zustand s , ist $\alpha_{ip|s}$ die intrinsische Neigung zum Kauf und $\beta_{1,ip|s}$ und $\beta_{2,ip|s}$ sind die entsprechenden Antwortparameter. Beachten Sie, dass wir im Gegensatz zum CEOM keine Identifikationsbeschränkungen für die Parameter des CPM auferlegen. Mit anderen Worten, wir machen keine Einschränkungen wie $\alpha_{ip|1} \leq \alpha_{ip|2} \leq \dots \leq \alpha_{ip|NS}$ und/oder $\delta_{ip|1} \leq \delta_{ip|2} \leq \dots \leq \delta_{ip|NS}$. Diese freiheitsgradreiche Spezifikation (im CPM) ermöglicht es unserem Modell flexibel zu sein, sodass beispielsweise Kunden in einem aktiveren E-Mail-Öffnungsstatus weniger aktiv bei Käufen sein können oder umgekehrt. Mit anderen Worten, anstatt Beschränkungen vorzugeben, dass E-Mail-aktive Kunden bei Käufen aktiv sein müssen, lassen wir die Daten zeigen, ob dies wirklich der Fall ist.

Um die unerkannte Heterogenität zu berücksichtigen, erlauben wir $\alpha_{ip|s}$, $\beta_{1,ip|s}$ und $\beta_{2,ip|s}$ kundenspezifisch zu sein. Wir gehen davon aus, dass $\alpha_{ip|s}$ und $\beta_{k,ip|s}$ (für $k = 1, 2$) bei den Kunden wie folgt normal verteilt sind:

$$\begin{aligned}\alpha_{ip|s} &= \alpha_{p|s} + \nabla \alpha_{ip} \text{ und} \\ \beta_{k,ip|s} &= \beta_{k,p|s} + \nabla \beta_{k,ip}, k = 1, 2, , \\ \text{wobei } \nabla \alpha_{ip} &\sim N(0, \sigma_{\alpha_p}^2) \text{ und } \nabla \beta_{k,ip} \sim N(0, \sigma_{\beta_{k,p}}^2) \text{ für } k = 1, 2\end{aligned}\tag{11}$$

5.5 Die Korrelation zwischen Kaufentscheidung und E-Mail-Öffnungen

In jedem Monat können sowohl die Anzahl der geöffneten E-Mails als auch die getätigten Käufe durch einen Kunden dessen Interesse und Interaktionsniveau mit dem Unternehmen anzeigen. Daher argumentieren wir, dass es eine mögliche Korrelation zwischen den Verteilungen der Kauf- und E-Mail-Öffnungszahlen gibt. Entsprechend dem Satz von Sklar (1959) können kumulative Verteilungsfunktionen beliebiger Variablen mithilfe einer Copula-Funktion verbunden werden, und diese Copula-Funktion ist eindeutig, wenn die beiden Variablen kontinuierlich sind. Bei zwei kontinuierlichen Variablen kann die bivariate Dichte aus den partiellen Ableitungen der gewählten Copula-Funktion abgeleitet werden. Bei zwei diskreten Variablen, wie den bivariaten Zähldaten in dieser Arbeit, können wir uns zwar nicht auf partielle Ableitungen verlassen, aber wir können trotzdem die bivariate Wahrscheinlichkeitsmassefunktion mithilfe von endlichen Differenzen der gewählten Copula-Funktion erhalten. Wir verwenden eine Copula, um die bivariate Wahrscheinlichkeitsmassefunktion von O_{it} und Y_{it} wie folgt zu konstruieren:

$$\begin{aligned}h(o_{it}, y_{it}) &= C[F_1(o_{it}), F_2(y_{it})] - C[F_1(o_{it} - 1), F_2(y_{it})] - C[F_1(o_{it}), F_2(y_{it} - 1)] \\ &\quad + C[F_1(o_{it} - 1), F_2(y_{it} - 1)]\end{aligned}\tag{12}$$

wobei $C(\cdot)$ die Copula-Funktion ist und $F_1(o_{it})$ und $F_2(y_{it})$ die Verteilungsfunktionen von O_{it} und Y_{it} sind. In diesem Zusammenhang verwenden wir eine Frank-Copula-Funktion aufgrund ihrer Flexibilität, den gesamten Bereich der Korrelation zu erfassen. Die Frank-Copula-Funktion ist gegeben durch:

$$C(u_1, u_2; \theta) = -\frac{1}{\theta} \log \left[1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right]\tag{13}$$

Hierbei sind u_1 und u_2 die Verteilungsfunktionen (ZINBD und BD in unserem Fall) und ρ ist der Korrelationsparameter der Frank-Copula.

5.6 Modellbestimmung

Es gibt vier Parameter-Sets, die aus unserem Modell geschätzt werden müssen: (1) Parameter der Übergangsmatrix, (2) Parameter der CEOM, (3) Parameter der CPM und (4) der Korrelationsparameter der Frank-Kopula-Funktion. Wir schreiben den Vektor der bivariaten Wahrscheinlichkeitsmassefunktion als Diagonalmatrix H_{it} . Gegeben der vorgeschlagenen HMM-Struktur kann die Likelihood-Funktion für eine Sequenz von Beobachtungen $[(Y_{i1} = y_{i1}, O_{i1} = o_{i1}), \dots, (Y_{it} = y_{it}, O_{it} = o_{it})]$ ausgedrückt werden:

$$L = \prod_{i=1}^N P[(Y_{i1} = y_{i1}, O_{i1} = o_{i1}), \dots, (Y_{it} = y_{it}, O_{it} = o_{it})] = \prod_{i=1}^N \psi_i' H_{i1} \prod_{t=2}^T Q_{it} H_{it} \mathbf{1} \quad (14)$$

wobei $\mathbf{1}$ ein $S \times 1$ Vektor aus Einsen ist.

6 Auswertung

6.1 Datenauswahl

Grundsätzlich betrachten wir Daten von Oktober 2021 bis Dezember 2022. Abbildung 8 verdeutlicht, dass eine Betrachtung der Daten vor Oktober 2021 nicht repräsentativ für die Kunden des Thalia Theaters wäre, da durch Beschränkungen während der Corona Pandemie eine kaum repräsentative Anzahl an Ticketverkäufen erfolgt ist. Dennoch ist festzustellen, dass nach Oktober 2021 ein ähnliches Niveau der Verkaufszahlen zu den Verkaufszahlen vor Corona erreicht wurde, siehe Abbildung 8.

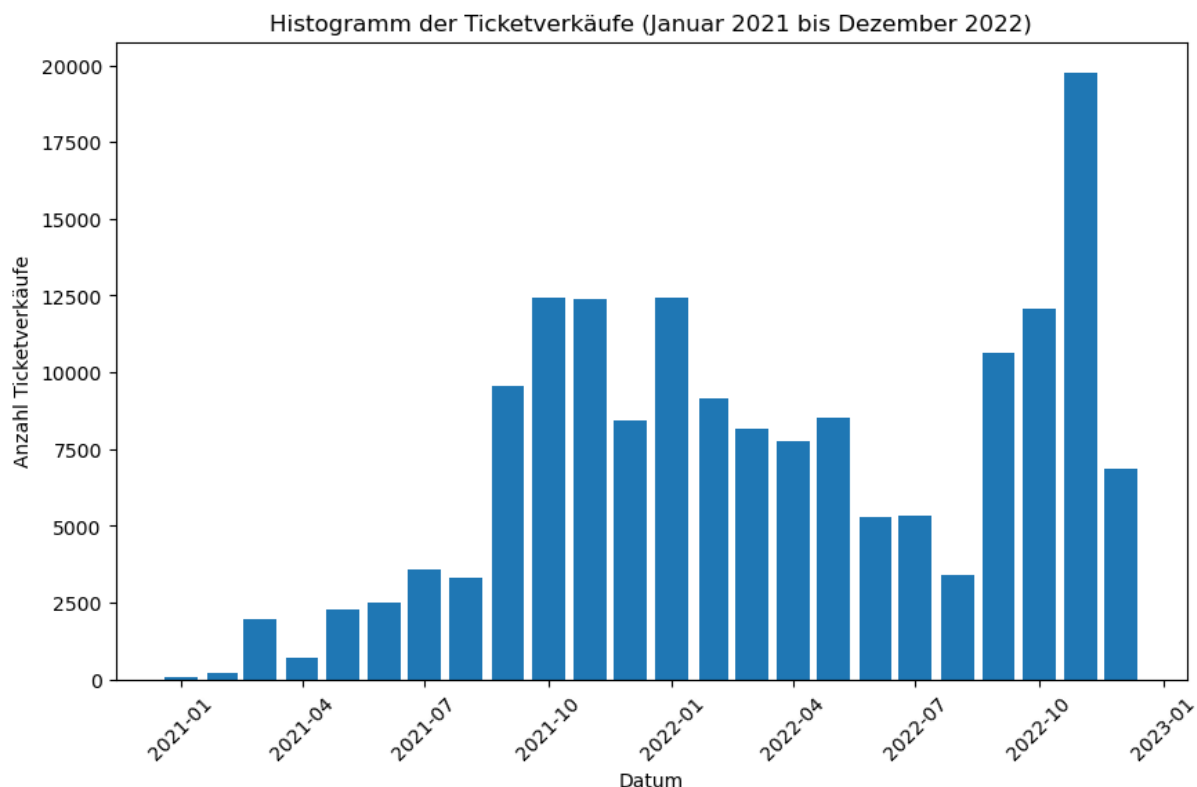


Abbildung 8: Histogramm der Ticketverkäufe (Januar 2021 bis Dezember 2022)

Für die Berechnungen sind sowohl die Ticketverkäufe, Messung erhaltener Newsletter-E-Mails, als auch die E-Mail-Öffnungen relevant. Im Folgenden wird erläutert, wie diese vor den Berechnungen bereinigt wurden. Zuvor wird beschrieben, wodurch eine Einschränkung auf Privatkunden durchgeführt wurde.

Privatkunden

Es ist leicht verständlich, dass eine Messung der Auswirkungen der E-Mail-Kampagnen auf die Kundenbindung eine Einschränkung auf Privatkunden erfordert. Um diese Einschränkung durchführen zu können, bedarf es einer Metrik, die Privatkunden von Geschäftskunden unter-

scheiden kann. Dabei hat sich die Größe "durchschnittliche Karten pro Bestellung" als geeignet herausgestellt. Es wurde das 1,5-fache des Interquartilabstands plus Quartil drei als obere Grenze, ein durchschnittlicher Wert von 3.25 Tickets pro Bestellung, festgelegt. Nach Bereinigung der Transaktionsdaten nach diesem Vorgehen kann noch über 71% des Gesamtumsatzes abgebildet werden. Des Weiteren sind über die Hälfte der 30 größten Privatkunden und keiner der 30 größten Geschäftskunden nach händischer Überprüfung noch in dem Datenpool enthalten.

Transaktionsdaten

In dieser Arbeit werden Transaktionen mit mehreren Tickets innerhalb einer Transaktion als eine einzige Transaktion gewertet. Wir betrachten eine Transaktion als eine Interaktion des Privatkunden mit dem Thalia Theater. Wenn die Größe der Transaktion in Betracht gezogen werden würde, wären mehrere Interaktionen des Kunden mit dem Theater innerhalb eines Besuchs möglich. Des Weiteren werden ausschließlich Kunden betrachtet, die im Oktober 2021 eine Transaktion getätigt haben und nach 6 Monaten mindestens eine weitere. Diese Einschränkung hat den Hintergrund, dass der Übergang von der Kundenloyalität zur Kundenbindung das Untersuchungskriterium ist. Indem ausschließlich Kunden ausgewählt werden mit einer zusätzlichen Transaktion nach 6 Monaten, wird sichergestellt, dass z.B. Touristen und Kunden, die grundsätzlich unzufrieden mit der Leistung des Theaters waren, ausgeschlossen werden.

Messung erhaltener Newsletter-E-Mails und E-Mail-Öffnungen

Bei der Auswahl der E-Mail-Kampagnen für die Berechnung findet eine Einschränkung auf Newsletter statt. Kunden erhalten sowohl eine automatische E-Mail zur Bestätigung der Newsletter-Anmeldung, eine Willkommensmail nach der Anmeldung, als auch zusätzliche Informationen nach einer Transaktion per E-Mail. Für die Berechnung des Modells werden ausschließlich E-Mails verwendet, die als "Newsletter" markiert sind. Wie in Kapitel **Fehler! Verweisquelle konnte nicht gefunden werden.** beschrieben, sind gemessene E-Mail-Öffnungen der E-Mail-Programme Apple Mail und Gmail als kritisch zu betrachten. Deshalb wurden die Daten derjenigen Kunden, bei denen festgestellt wurde, dass sie mindestens eine E-Mail über diese Programme geöffnet haben, nicht verwendet.

6.2 Modellidentifikation

Die Parameter des HMM werden mithilfe der Maximum-Likelihood-Schätzmethode geschätzt. Dabei wird die absolute logarithmierte Wahrscheinlichkeit anhand der in 5.1 beschriebenen Daten und Gleichung (15) minimiert. Bei der Anwendung ist festzustellen, dass lediglich eine Anzahl von zwei Zuständen für die verwendeten Daten sinnvoll ist. Eine Auswertung mit drei oder mehr Zuständen hat zwar ähnliche Wahrscheinlichkeiten erzielen können, aber hat immer

einen der drei Zustände für sehr spezifische Kunden festgelegt und damit keine repräsentativen Ergebnisse erzielt. Daher wird im Folgenden das HMM mit zwei Zuständen für weitere Analysen ausgewählt.

BFGS-Algorithmus zur Berechnung der Wahrscheinlichkeiten

Bei der Minimierung der absoluten logarithmierten Wahrscheinlichkeit zur Parameterschätzung wird der BFG-Algorithmus verwendet. Der Broyden-Fletcher-Goldfarb-Shanno-Algorithmus (BFGS-Algorithmus) ist ein lokaler Such-Optimierungsalgorithmus. Er gehört zur Klasse der sogenannten Quasi-Newton-Methoden, da er eine Näherung der Hesse-Matrix, also der zweiten Ableitung einer Zielfunktion, verwendet. Der BFGS-Algorithmus ist ein Optimierungsalgorithmus zweiter Ordnung, der für Optimierungsprobleme eingesetzt wird, bei denen die zweite Ableitung nicht berechnet werden kann.

Startparameter

Vor den Berechnungen wird zunächst der Dispersion-Parameter r der ZINBD-Verteilung geschätzt. Ein größerer r -Wert führt zu einer geringeren Überdispersion, d.h. einer geringeren Streuung der Daten, und einer stärkeren Konzentration der Daten um den Erwartungswert. Für die Schätzung von r wurde die empirische Verteilung der Käufe pro Monat herangezogen.

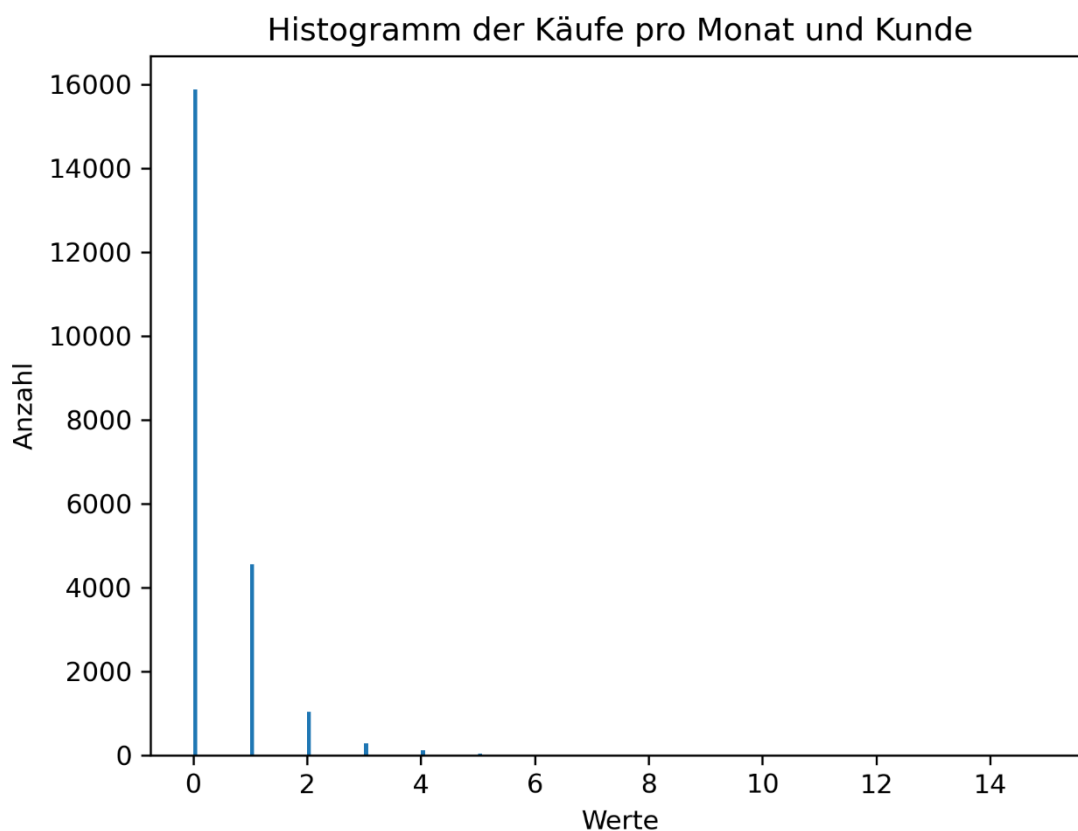


Abbildung 9: Histogramm der Käufe pro Monat und Kunde (im Zeitraum Oktober 2021 bis Dezember 2022)

Abbildung 9: Histogramm der Käufe pro Monat und Kunde (im Zeitraum Oktober 2021 bis Dezember 2022) zeigt das Histogramm der Anzahl an Käufen pro Monat und Kunde. Aus diesen Daten wird die empirische Wahrscheinlichkeitsfunktion der Käufe ermittelt. Es werden die quadratischen Fehler der Wahrscheinlichkeitsfunktion einer ZINBD-, NBD und Poisson-verteilte Zufallsvariablen gegenüber der empirischen Wahrscheinlichkeitsfunktion minimiert. Für die ZINBD-Verteilung ergibt sich ein geschätzter Wert von 1.29 für den Dispersion-Parameter r . Abbildung 10: Vergleich der Verteilungsfunktionen (CDF) einer ZINBD-, NBD- und Poisson-verteilten ZV mit geschätzten Parameterwerten gegenüber der empirischen Verteilungsfunktion

Abbildung 10 zeigt die Verteilungsfunktionen mit den geschätzten Parameterwerten gegenüber der empirischen Verteilungsfunktion. Die Verteilungsfunktion der ZINBD-verteilten Zufallsvariable nimmt dabei dieselben Werte der NBD-verteilten Zufallsvariable an, da die Zero-Inflation-Komponente auf den Wert 0 geschätzt wird. Dieser Umstand legt nahe, dass eine NBD-Verteilung anstatt einer ZINBD-Verteilung für die Modellierung verwendet wird. Da die Definition des Modells jedoch so erfolgt, dass das Modell auch für zukünftige Berechnungen in der Praxis für Neukunden und andere Kunden mit größeren Intervallen zwischen Käufen anwendbar sein soll, wird die Definition möglichst allgemein gehalten und die Zero-Inflation-Komponente weiterhin berücksichtigt.

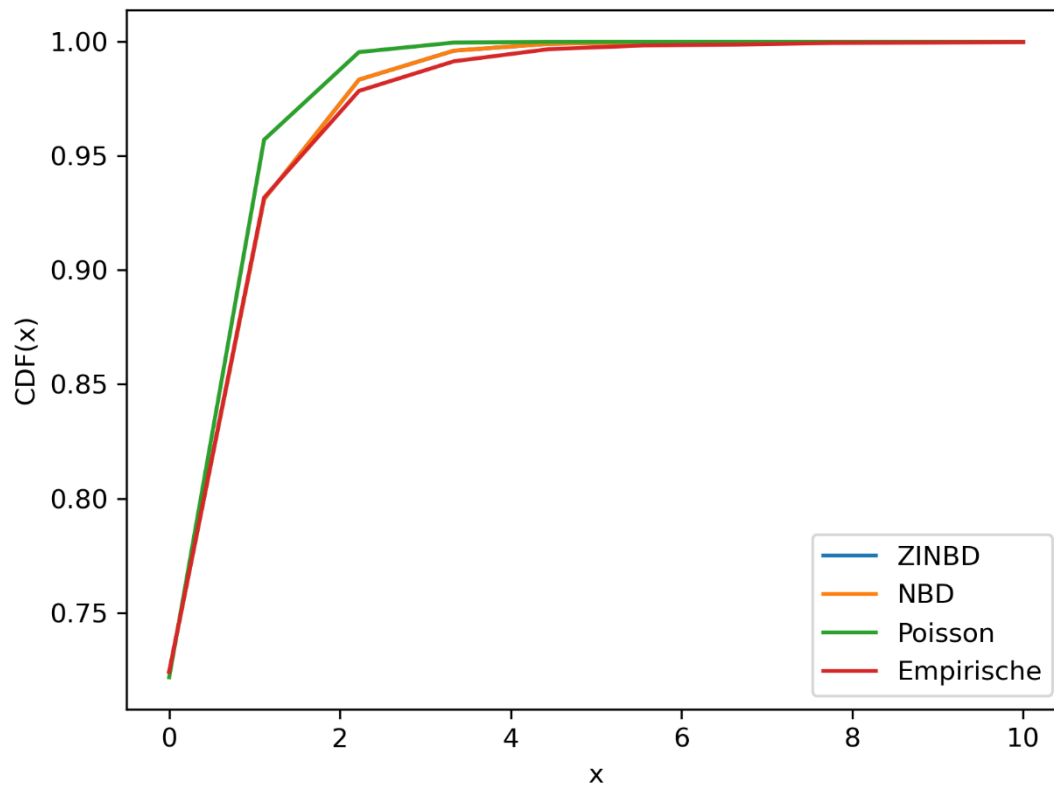


Abbildung 10: Vergleich der Verteilungsfunktionen (CDF) einer ZINBD-, NBD- und Poissonverteilten ZV mit geschätzten Parameterwerten gegenüber der empirischen Verteilungsfunktion

Da das Modell dazu genutzt wird, die Einflüsse des E-Mail-Marketings auf die Kundenbindung selbst zu erkennen und keine vergleichbaren Ergebnisse zum Einfluss der Parameter vorliegen, starten alle anderen Parameterwerte mit dem Wert 0.

Parameter	Startwerte	Untere Grenze	Obere Grenze
alpha	[[0, 0], [0, 0]]	-3	3
gamma_1	[[0, 0], [0, 0]]	-3	3
gamma_2	[[0, 0], [0, 0]]	-3	3
gamma_3	[[0, 0], [0, 0]]	-3	3
gamma_4	[[0, 0], [0, 0]]	-3	3
alpha_o	[0, 0]	-3	3
beta_o	[0]	-3	3
delta_0	[0, 0]	-8	8
delta_1	[0, 0]	-3	2

alpha_p	[0, 0]	-3	3
beta_p_1	[0, 0]	-3	3
beta_p_2	[0, 0]	-3	3
r (fest)	[1.29]		
theta	[0]	0.1	keine

Tabelle 1: Startparameter und Grenzen (für die einzelnen Werte in den Parametern)

Grenzen für die Parameter

Es hat sich in der praktischen Anwendung des Modells bewährt, Grenzen für die Parameter zu bestimmen, um unrealistische Werte zu vermeiden und die numerische Stabilität zu gewährleisten. In Tabelle 1 werden die oberen und unteren Grenzen für die Parameter festgelegt.

6.3 Parameterbestimmungen

Im folgenden Abschnitt werden zunächst die Parameterschätzungen des vorgeschlagenen HMMs und des Copula-Modells für das Kauf- und E-Mail-Öffnungsverhalten der Kunden diskutiert. Anschließend wird verdeutlicht, wie die E-Mail-Kontaktentscheidung des Unternehmens die Beziehungszustände zwischen Kunden und Unternehmen beeinflusst. Insbesondere wird erläutert, wie diese Entscheidung die Übergänge der Kunden zwischen verschiedenen Beziehungszuständen beeinflusst. Des Weiteren wird die Verteilung der Kunden in den zwei Kunden-Unternehmens-Beziehungszuständen diskutiert. Es wird der Anteil der Kunden im Laufe der Zeit dargestellt, die verschiedenen Beziehungszuständen zugeordnet sind.

Parameterschätzungen

In **Fehler! Verweisquelle konnte nicht gefunden werden.** sind die Parameterabschätzungen für das HMM mit zwei Zuständen dargestellt. Zur Charakterisierung der Zustände hinsichtlich des Kauf- und E-Mail-Öffnungsverhaltens werden die erwarteten Kauf- und E-Mail-Öffnungszahlen berechnet (basierend auf beobachteten Werten der Verhaltensvariablen auf Kundenebene und Unternehmensebene wie LY, LO, EM usw.). Es wird festgestellt, dass Kunden im Zustand 1 durchschnittlich 0.485 E-Mails pro Monat öffnen, während Kunden im zweiten Zustand 1.405 E-Mails pro Monat öffnen. Bezüglich des Kaufverhaltens wird festgestellt, dass Kunden im Zustand 1 durchschnittlich 0.234 Käufe pro Monat tätigen, während Kunden im Zustand 2 0.159 Käufe pro Monat tätigen. Basierend auf dieser Berechnung werden die beiden latenten Zustände, die die Häufigkeit der E-Mail-Öffnung und des Kaufs bei Kunden in jedem Monat berücksichtigen, als "wenige Öffnungen / viele Käufe" und "viele Öffnungen / mittlere Käufe" bezeichnet. Die Ergebnisse widerlegen die herkömmliche Weisheit, dass Kunden, die die meisten Käufe tätigen, auch die meisten E-Mails öffnen.

Interpretation der Parameter für die CPM

Parameter	Zustand 1	Zustand 2
delta_0	7.701	8.000
delta_1	-1.209	1.995
alpha_p	-1.110	-1.893
beta_p_1	0.015	0.179
beta_p_2	-0.039	-0.035
r (zustandsinvariant)	1.29	

Tabelle 2: Relevante Parameter für die Modellierung des Kaufverhaltens

Die Anzahl der Kundenkäufe wurde mittels einer ZINBD-Verteilung modelliert, wobei die Parameter δ_0 und δ_1 den Wert von $\phi_{it|s}$ in Gleichung (8) beeinflussen. $\phi_{it|s}$ ist ein Parameter, der die Wahrscheinlichkeit von Null-Ereignissen in der Verteilung steuert. Er repräsentiert den Anteil beobachteter Null-Werte im Vergleich zu erwarteten Null-Werten. Genauer gesagt gibt $\phi_{it|s}$ den Anteil der Null-Ereignisse an, der nicht auf die Negative Binomial Distribution (NBD) zurückzuführen ist. Gemäß der Definition von Gleichung (8) führen höhere Werte von δ_0 und δ_1 zu einem kleineren $\phi_{it|s}$.

Nun wird der konkrete Einfluss der Parameter auf die Verteilung der Anzahl der Kundenkäufe betrachtet. Der beobachtete Wert LY_{it} , der die Anzahl der Monate seit dem letzten Kauf darstellt, nimmt bei den verwendeten Daten Werte zwischen 1 und 14 an. Man erkennt, dass die Zero-Inflation-Komponente der Verteilung für den Zustand 2 kaum eine Rolle spielt, da $\phi_{it|2}$ Werte kleiner als 0.001 für $LY_{it} = 1, \dots, 14$ annimmt.

In Zustand 1 wiederum kann für größere Werte von LY_{it} , das heißt ein größerer Zeitraum seit dem letzten Kauf, ein zunehmender Einfluss der Zero-Inflation-Komponente beobachtet werden. Dennoch kann auch im ersten Zustand der Einfluss der Zero-Inflation-Komponente vernachlässigt werden. So nimmt $\phi_{it|1}$ Werte im Intervall $[0.0001, 0.0109]$ für $LY_{it} = 1, \dots, 14$ an. Der Einfluss der Parameter α_p , β_{p_1} und β_{p_2} auf die Verteilung der Käufe ist in Gleichung (10) definiert. Bei den Berechnungen wurde r als 1.29 aufgrund von vorherigen Schätzungen festgelegt. Des Weiteren wurde gerade festgestellt, dass $\phi_{it|s}$ vernachlässigbar kleine Werte annimmt. **Fehler! Verweisquelle konnte nicht gefunden werden.** zeigt den Einfluss von λ auf die Wahrscheinlichkeitsfunktion der Käufe bei $r = 1.29$ und $\phi_{it|s} = 0$ für $i = 1, \dots, N$, $t = 1, \dots, T$ und $s = 1, 2, 3$.

Wahrscheinlichkeitsfunktion der Käufe für verschiedene Werte von lambda

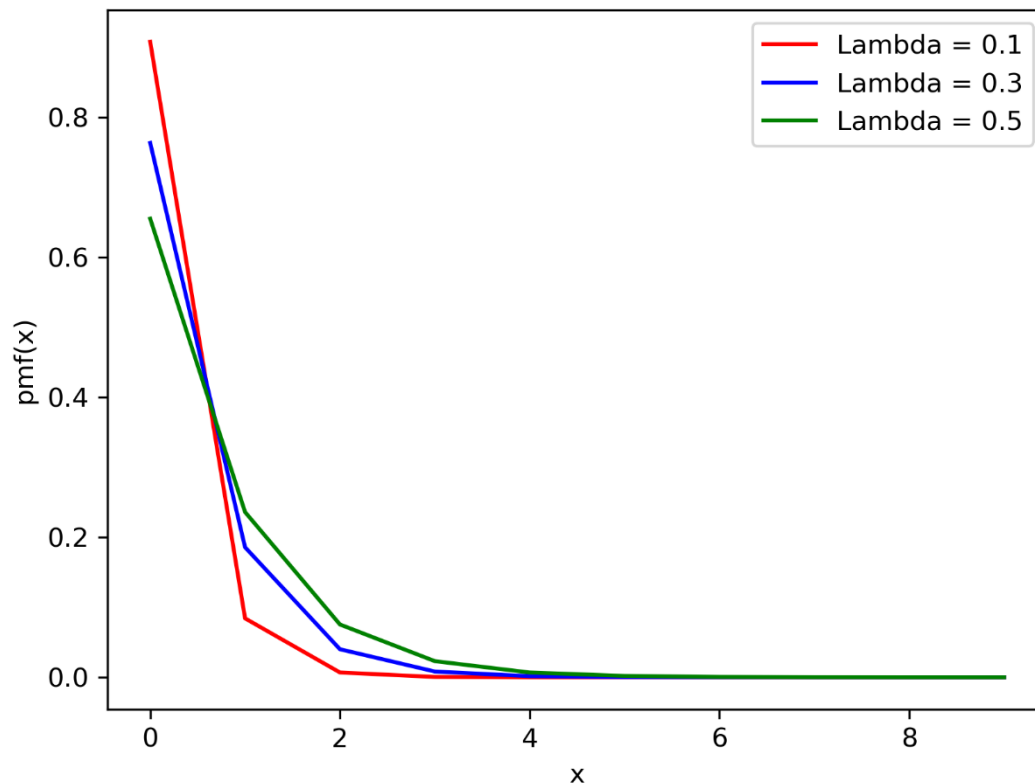


Abbildung 11: Wahrscheinlichkeitsfunktion der Käufe für verschiedene Werte von lambda

Im Betrachtungszeitraum erhielten Kunden durchschnittlich 1.54 E-Mails pro Monat, wobei das Maximum drei E-Mails in einem Monat beträgt. Durch Einsetzen der geschätzten Parameterwerte in Gleichung (10) und Verwendung der 1.54 E-Mails pro Monat, nimmt in einem durchschnittlichen Monat lambda für Kunden im Zustand 1 den Wert 0.308 und für Kunden in Zustand 2 den Wert 0.182 an. Mit $\varphi_{it|s} \approx 0$ kann lambda auch als der Erwartungswert für die Anzahl der Käufe interpretiert werden.

E-Mails in diesem Monat	lambda, Zustand 1	lambda, Zustand 2
0	0.329	0.150
1	0.321	0.174
2	0.291	0.186
3	0.243	0.187

Tabelle 3: Einfluss der E-Mail-Kampagnen auf das Kaufverhalten

Tabelle 3 zeigt, dass Kunden in Zustand 1 weniger E-Mails bevorzugen. Da β_{p_2} für diese Kundengruppe negativ ist, ist der negative Einfluss quadratisch. Es gilt, je mehr E-Mails an Kundengruppe 1 geschickt werden, desto unwahrscheinlicher ist deren Kauf. Kunden in Zustand 2 wiederum bevorzugen es grundsätzlich mehr E-Mails zu erhalten, wobei der Einfluss des zusätzlichen E-Mail-Erhalts einen geringeren Einfluss auf deren Kaufverhalten hat.

Interpretation der Parameter für die CEOM

In Tabelle 4: Relevante Parameter für die Modellierung der E-Mail-Öffnungen sind die Schätzungen der relevanten Parameter für die Modellierung der E-Mail-Öffnungen dargestellt. Alle Parameter fließen in die Berechnung des Parameters λ der Poissonverteilung ein. Somit haben alle Parameter einen direkten und mit der Definition in (5) exponentiellen Einfluss auf die erwartete Anzahl an Öffnungen. Es besteht ein großer Unterschied im E-Mail-Öffnungsverhalten der beiden Kundengruppen. So öffnen Kunden im Zustand 2 im Durchschnitt $\exp(1.396 - 0.331) \approx 2.9$ mal mehr E-Mails als Kunden im Zustand 1. β_{t_O} spiegelt den Einfluss der Zeit seit der letzten E-Mail-Öffnung wider. Pro vergangenem Monat seit der letzten Öffnung sinkt die erwartete Anzahl der E-Mail-Öffnungen um ca. 32.5%.

Parameter	Zustand 1	Zustand 2
Alpha_O	0.331	1.396
Beta_O (zustandsinvariant)	-1.121	

Tabelle 4: Relevante Parameter für die Modellierung der E-Mail-Öffnungen

Übergangsmatrix

Die Übergangsmatrix des HMM zeigt, wie Kunden sich in verschiedenen Beziehungszuständen entwickeln. Wir berechnen die Übergangswahrscheinlichkeiten eines "typischen" Kunden mithilfe der Gleichungen 1-4. Der „typische“ Kunde hatte eine Öffnung im letzten Monat getätigt, da dieses Kriterium im Schnitt auf 58,00% der Kunden im Zeitverlauf zutrifft. Des Weiteren hatte der Kunde keinen Kauf im letzten Monat, da dies für einen zufällig ausgewählten Kunden mit einer Wahrscheinlichkeit von 62,00% zutrifft. Wir variieren die Anzahl der in der vorherigen Periode erhaltenen E-Mails und überprüfen die Auswirkungen von E-Mail-Kontakten auf die Zustandsübergänge (siehe Tabelle 5).

	Zu Zustand 1 („viele Käufe“)	Zu Zustand 2 („mittlere Käufe“)
Von Zustand 1		
<i>Keine erhaltene E-Mail</i>	30.13%	69.86%
<i>Eine erhaltene E-Mail</i>	9.53%	90.47%

<i>Zwei erhaltene E-Mails</i>	2.76%	97.24%
<i>Drei erhaltene E-Mails</i>	0.83%	99.16%
Von Zustand 2		
<i>Keine erhaltene E-Mail</i>	2.20%	97.80%
<i>Eine erhaltene E-Mail</i>	9.22%	90.78%
<i>Zwei erhaltene E-Mails</i>	11.68%	88.32%
<i>Drei erhaltene E-Mails</i>	4.74%	95.26%

Tabelle 5: Veranschaulichung der Übergangsmatrix anhand eines „typischen“ Kunden (eine Öffnung im letzten Monat und kein Kauf)

Tabelle 5 zeigt, dass E-Mail-Kontakt unterschiedliche Auswirkungen auf die Übergangswahrscheinlichkeiten von Kunden hat, die sich in verschiedenen Kaufzuständen befinden. Um die Wahrscheinlichkeit zu maximieren, dass Kunden in dem Zustand mit vielen Käufen bleiben, sollten keine E-Mails an diese Kundengruppe versendet werden. Um Kunden von dem Zustand der „mittleren Käufe“ in den Zustand der „vielen Käufe“ zu bewegen, sollten zwei E-Mails pro Monat an diese Kunden versendet werden.

	Zu Zustand 1 („viele Käufe“)	Zu Zustand 2 („mittlere Käufe“)
Von Zustand 1		
<i>Keine erhaltene E-Mail</i>	60.83%	39.17%
<i>Eine erhaltene E-Mail</i>	27.50%	72.50%
<i>Zwei erhaltene E-Mails</i>	9.27%	90.73%
<i>Drei erhaltene E-Mails</i>	2.94%	97.06%
Von Zustand 2		
<i>Keine erhaltene E-Mail</i>	1.10%	98.90%
<i>Eine erhaltene E-Mail</i>	4.81%	95.19%
<i>Zwei erhaltene E-Mails</i>	6.17%	93.83%
<i>Drei erhaltene E-Mails</i>	2.41%	97.59%

Tabelle 6: Veranschaulichung der Übergangsmatrix anhand eines Kunden mit einer Öffnung im letzten Monat und einem Kauf

Tabelle 6 zeigt die Übergangsmatrix anhand eines Kunden mit einer Öffnung im letzten Monat und einem Kauf. Im Vergleich zu der Übergangsmatrix eines „typischen“ Kunden ohne Kauf im vorherigen Monat ist auffallend, dass Kunden im Zustand „viele Käufe“ eine mehr als doppelt so hohe Wahrscheinlichkeit (60.83% im Vergleich zu 30.13%) haben, in ihrem Zustand zu

bleiben. Dennoch hat die Anzahl der versendeten Mails an diese Kundengruppe einen erheblichen Einfluss auf deren Verhalten im Modell. So kann die Wahrscheinlichkeit, dass sie den Zustand verlassen, bei drei erhaltenen Mails auf 97.06% erhöht werden.

Des Weiteren ist der Einfluss des E-Mail-Erhalts auf die Übergangswahrscheinlichkeiten von Kunden im Zustand „mittlere Käufe“ deutlich geringer, falls sie im letzten Monat einen Kauf getätigt haben. Dies lässt sich daran erkennen, dass deren Wahrscheinlichkeit im Zustand 2 zu bleiben für alle Anzahlen an versendeten Mails größer ist als bei den „typischen“ Kunden. Dennoch gilt auch hier, dass zwei E-Mails die optimale Anzahl an E-Mails ist, um Kunden vom Zustand der „mittleren Käufe“ in den Zustand der „hohen Käufe“ zu bewegen.

Grundsätzlich lässt sich feststellen, dass ein Kauf im letzten Monat dazu führt, dass Kunden eher dazu tendieren, in ihrem Zustand zu bleiben.

	Zu Zustand 1 („viele Käufe“)	Zu Zustand 2 („mittlere Käufe“)
Von Zustand 1		
<i>Keine erhaltene E-Mail</i>	86.31%	13.69%
<i>Eine erhaltene E-Mail</i>	60.62%	39.28%
<i>Zwei erhaltene E-Mails</i>	29.31%	70.69%
<i>Drei erhaltene E-Mails</i>	10.97%	89.03%
Von Zustand 2		
<i>Keine erhaltene E-Mail</i>	27.09%	72.01%
<i>Eine erhaltene E-Mail</i>	62.68%	37.32%
<i>Zwei erhaltene E-Mails</i>	68.64%	31.36%
<i>Drei erhaltene E-Mails</i>	45.13%	54.87%

Tabelle 7: Veranschaulichung der Übergangsmatrix anhand eines Kunden mit keiner Öffnung im letzten Monat und keinem Kauf

Bei der Analyse der Übergangsmatrix von Kunden ohne Öffnung und Kauf im letzten Monat fällt auf, dass Kunden im Zustand „mittlere Käufe“ sehr viel eher dazu neigen, danach in den Zustand der „vielen Käufe“ zu wechseln, im Vergleich zu den vorherigen Betrachtungen. Kunden, die bereits in Zustand 1 sind, neigen ebenfalls eher dazu, in dem Zustand zu bleiben. Zuletzt werden noch die „untypischsten“ Kunden betrachtet – Kunden, bei denen keine Öffnung im letzten Monat gemessen wurde, die aber einen Kauf getätigt haben. Es ist zu erkennen, dass für diesen Fall die Wahrscheinlichkeit für Kunden in Zustand 1 in diesem Zustand zu bleiben, maximiert wird. So haben diese Kunden beim Nicht-Erhalt von E-Mails eine 95.78% Wahrscheinlichkeit in ihrem Zustand zu bleiben. Ebenfalls können Kunden in Zustand 2 beim

Versenden von zwei E-Mails an diese Kundengruppe mit einer über 50% Erfolgswahrscheinlichkeit dazu bewogen werden, in den Zustand „viele Käufe“ zu wechseln.

	Zu Zustand 1 („viele Käufe“)	Zu Zustand 2 („mittlere Käufe“)
Von Zustand 1		
<i>Keine erhaltene E-Mail</i>	95.78%	4.22%
<i>Eine erhaltene E-Mail</i>	84.72%	15.28%
<i>Zwei erhaltene E-Mails</i>	59.89%	40.11%
<i>Drei erhaltene E-Mails</i>	30.72%	69.28%
Von Zustand 2		
<i>Keine erhaltene E-Mail</i>	15.60%	84.44%
<i>Eine erhaltene E-Mail</i>	45.52%	54.58%
<i>Zwei erhaltene E-Mails</i>	52.12%	47.88%
<i>Drei erhaltene E-Mails</i>	29.03%	70.67%

Tabelle 9: Veranschaulichung der Übergangsmatrix anhand eines Kunden mit keiner Öffnung im letzten Monat und einem Kauf

6.4 Bestimmung des Kundenzustands

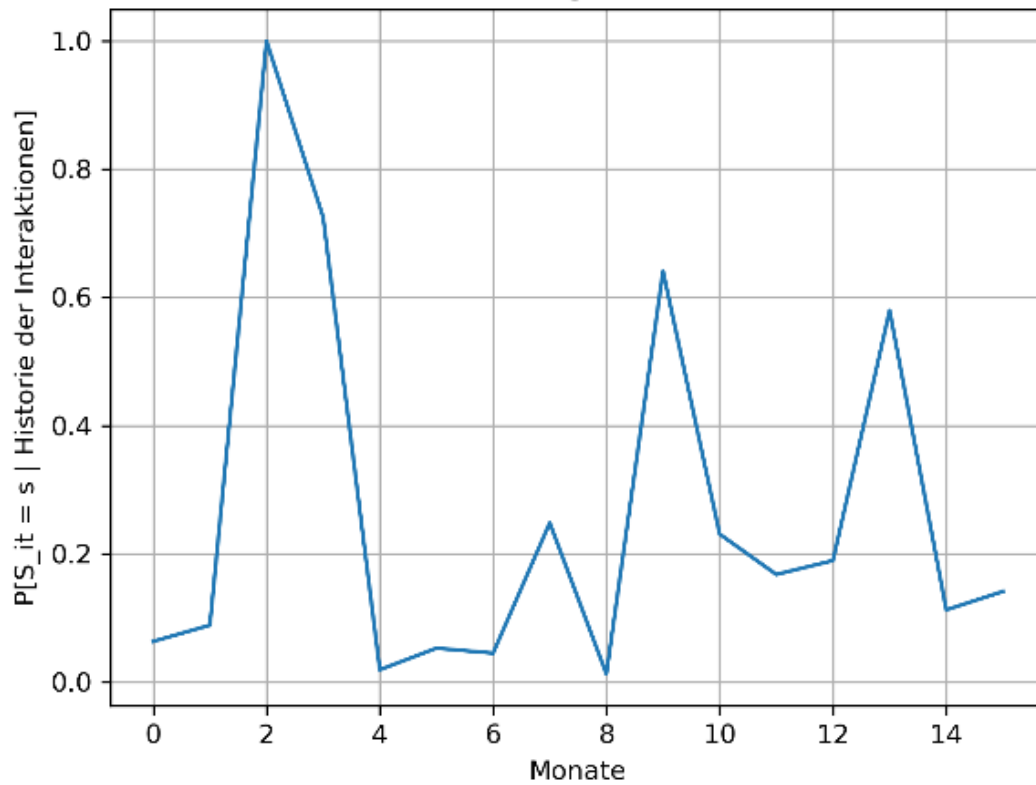
Ein attraktives Merkmal eines HMM ist die Fähigkeit, den Zustand einer Person zu jedem Zeitpunkt probabilistisch darzustellen. Diese Messung kann direkt aus der Wahrscheinlichkeitsfunktion in Gleichung (15) abgeleitet werden. Es wird bei Berechnung der Wahrscheinlichkeiten zwischen der Methode der Filterung und der Glättung unterschieden. Die Filterung verwendet nur die bis zum Zeitpunkt t bekannten Informationen, um den Zustand der Person zum Zeitpunkt t darzustellen, während die Glättung die vollständigen Informationen aus den Daten verwendet. Der Filterungsansatz ist für Marketinganwendungen attraktiver, bei welchen Entscheidungen nur auf der Historie des beobachteten Verhaltens basieren (vgl. Netzer et al. 2008, S. 192). Die Filterungswahrscheinlichkeit, dass Person i zum Zeitpunkt t im Zustand s ist, unter der Bedingung der Historie der Entscheidungen der Person, ist gegeben durch:

$$\begin{aligned}
 P[S_{i,t} = s \mid (Y_{i1} = y_{i1}, O_{i1} = o_{i1}), \dots, (Y_{it} = y_{it}, O_{it} = o_{it})] \\
 = \frac{(\psi_i' H_{i1} \prod_{k=2}^{T-1} Q_{ik} H_{ik}) * (H)_{i,t,s}}{L_{i,t}}
 \end{aligned} \tag{15}$$

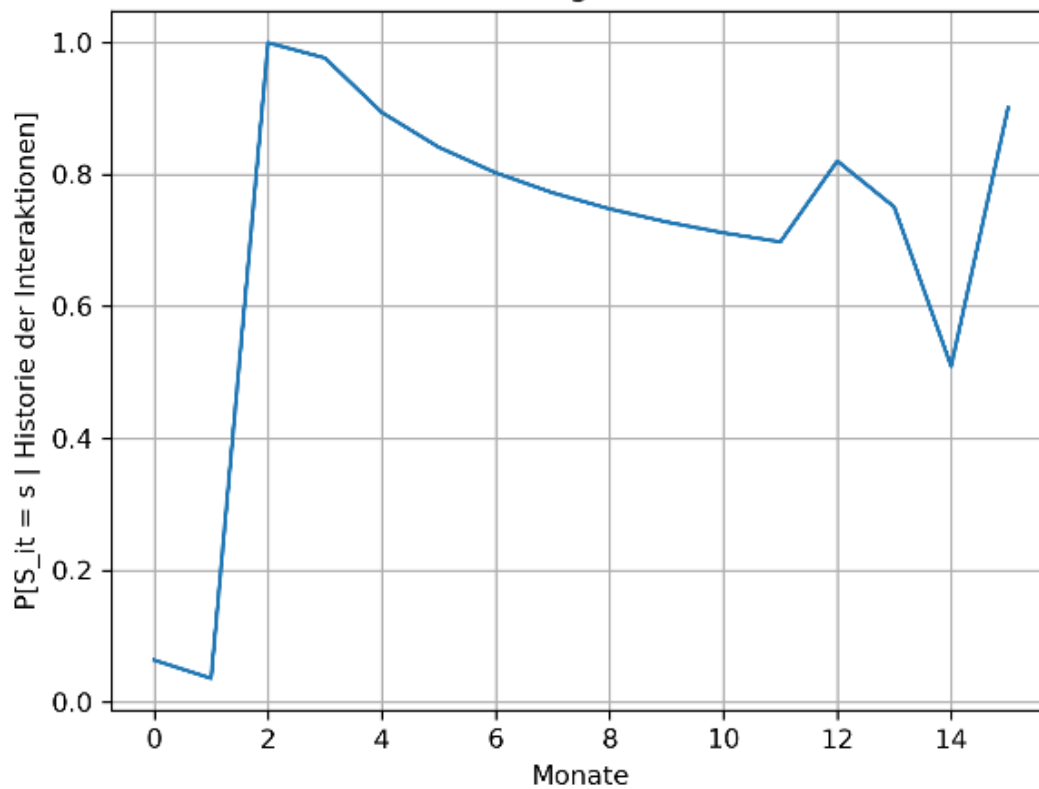
$L_{i,t}$ ist die Wahrscheinlichkeit der beobachteten Sequenz von Entscheidungen bis zum Zeitpunkt t gemäß Gleichung (14). $(H)_{i,t,s}$ gibt die s -te Spalte der Matrix an. In **Fehler! Verweisquelle konnte nicht gefunden werden.** sind die Wahrscheinlichkeiten, zufällig ausgewählter Kunden in Zustand 1 zu sein, abgebildet. Da alle Kunden nach dem Kriterium „Transaktion im Oktober 2021“ ausgewählt wurden, macht es Sinn, dass die Verteilungen sich in den ersten Monaten sehr ähneln. Deshalb sollten diese Ergebnisse erst ab ca. 5 Monaten interpretiert werden, da so die Unterschiede des Verhaltens der Kunden der entscheidende Faktor für deren Übergangswahrscheinlichkeit ist und nicht die getroffene Vorauswahl.

Es fällt auf, dass der Kundenbindungszustand große Schwankungen im Zeitverlauf aufweist. Selbst Kunden, die sehr stark an das Unternehmen gebunden sind (gemessen an einer Wahrscheinlichkeit in Zustand 1 zu sein von konstant über 50%), weisen teilweise starke Schwankungen in der Kundenbindung auf. Dies entspricht der Anfangsthese, dass die Kundenbindung sich über den Zeitverlauf ändert und es daher eines dynamischen Managements der Kundenbindung bedarf.

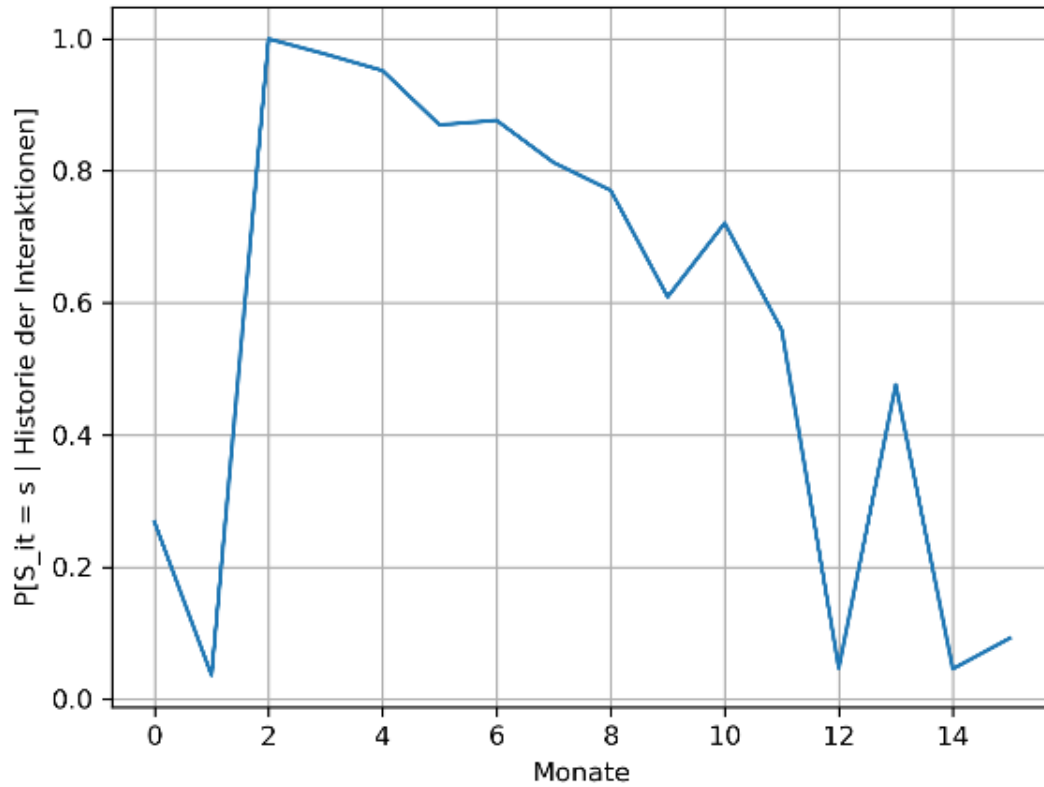
Wahrscheinlichkeit eines zufälligen Kunden in Zustand 1 zu sein



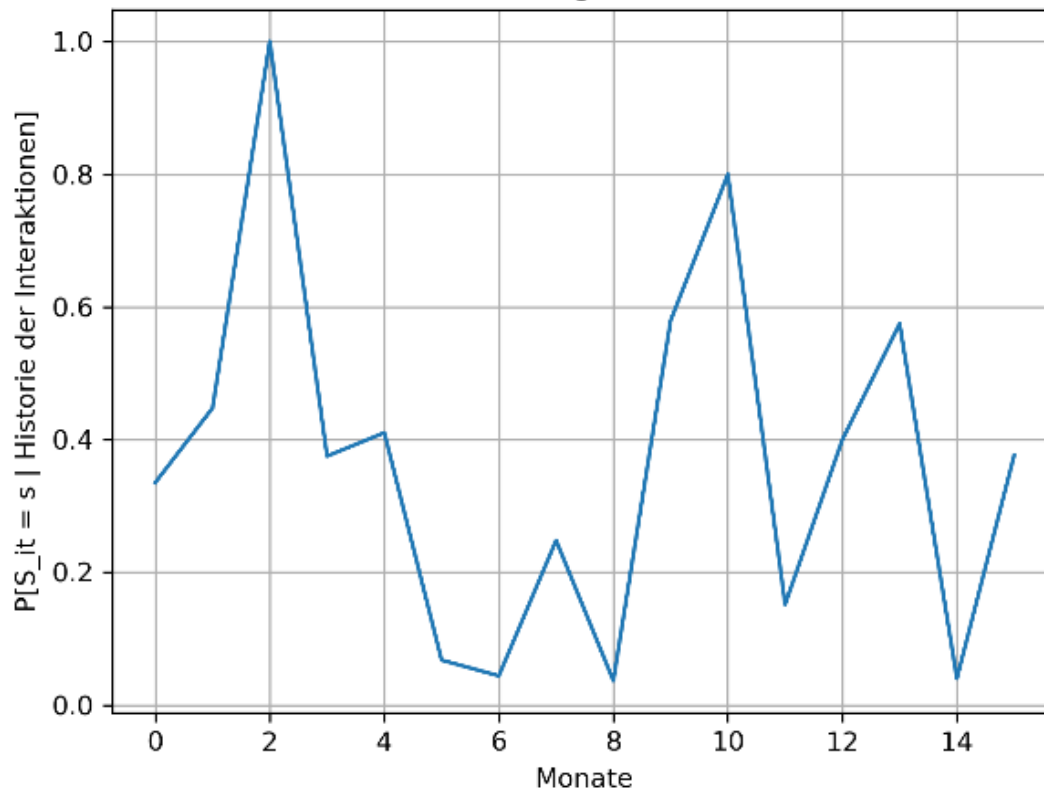
Wahrscheinlichkeit eines zufälligen Kunden in Zustand 1 zu sein



Wahrscheinlichkeit eines zufälligen Kunden in Zustand 1 zu sein



Wahrscheinlichkeit eines zufälligen Kunden in Zustand 1 zu sein



6.5 Zusammenfassung

Zusammenfassend hat der E-Mail-Kontakt abhängig vom Zustand der Kundenbindung einen variierenden Effekt auf die Bindung der Kunden. Grundsätzlich ist es das Ziel eines wirtschaftlich agierenden Unternehmens möglichst viele Kunden in den Zustand 1 der „vielen Käufe“ zu bewegen. Für Kunden aus dem Zustand mit hohen Käufen führen übermäßige E-Mail-Kontakte dazu, dass sie in Zustände mit geringeren Käufen abwandern. Während die Wahrscheinlichkeit, dass sie im Zustand mit hohen Käufen bleiben, im Szenario ohne E-Mails abhängig davon, ob sie im vorherigen Monat eine Öffnung und/oder Kauf getätigt haben, zwischen 30.13% und 95.78% beträgt, nimmt diese Wahrscheinlichkeit im Szenario mit drei E-Mails signifikant auf 0.83% bis 30.72% ab. Des Weiteren wurde ein linearer negativer Zusammenhang zwischen dem Kaufverhalten von Kunden mit hohen Käufen und den erhaltenen E-Mails festgestellt, siehe Tabelle 3. Für das Thalia Theater ist es daher empfehlenswert, keine E-Mails mehr an Kunden des Zustands 1 zu versenden.

Für Kunden in Zustand 2 wurde ein positiver Zusammenhang zwischen dem E-Mail-Erhalt und der Anzahl der Käufe aufgezeigt, siehe Tabelle 3. Bei der Betrachtung der Übergangsmatrizen ist aufgefallen, dass der gewünschte Effekt, Kunden von Zustand 2 in Zustand 1 zu bewegen, eine Art umgedrehte „U-Form“ mit einem Maximum bei zwei versendeten E-Mails aufweist. Da die erwartete Anzahl an Käufen der Kunden mit „mittleren Käufen“ bei einem E-Mail-Versand von drei statt zwei E-Mails lediglich von ca. 0.186 auf 0.187 steigt, aber eine deutliche Verringerung der Übergangswahrscheinlichkeit von Zustand 2 zu Zustand 1 zur Folge hat, ist es ratsam für das Thalia Theater, zwei E-Mails an diese Kundengruppe zu verschicken.

6.6 Robustheitstests

Um zu überprüfen, ob unser Modell und unsere Ergebnisse robust gegenüber alternativen Modellspezifikationen und auf andere Kunden anwendbar ist, führen wir drei Reihen von Robustheitstests durch:

1. Test mit Negativer Binomial-Verteilung zur Modellierung der Kaufzahlen
2. Tests mit alternativen Verteilungen zur Modellierung der Öffnungszahlen
3. Tests mit Kunden, die im November 2021 ihren ersten Kauf getätigt haben und nach 6 Monaten mindestens einen weiteren

NBD als alternative Verteilung zur Modellierung der Kaufzahlen

Als ersten Robustheitstest schätzen wir die alternative Modellspezifikationen, dass die Kaufanzahl als negative Binomialverteilung (NBD) verteilt ist. Es werden eine ähnliche Korrelation zwischen Käufen und Öffnungen festgestellt. Es werden ebenfalls die Kunden in ähnliche Kundengruppen segmentiert. Das Modell mit der NBD-Verteilung schätzt ebenfalls einen negativen Zusammenhang zwischen mehr als zwei E-Mail-Versendungen an Kunden in Zustand 1

und deren erwarteten Käufen. Es schätzt allerdings einen positiven Einfluss der Versendung von einer E-Mail pro Monat an diese Kunden. Trotz ähnlicher Segmentierungen und Schätzungen des Kundenverhaltens, wird ein schlechterer Wert für die absolute logarithmierte Wahrscheinlichkeit erzielt von 20927.46, im Vergleich zu 17936.24 mit dem in dieser Arbeit vorgestellten Modell. Abbildung 10 ist bereits ein starkes Indiz dafür, dass eine Poisson-Verteilung zur Modellierung der Kaufzahlen zu schlechteren Ergebnissen führen würde. Es wurde daher davon abgesehen, diese Verteilung ebenfalls zu testen. Des Weiteren erfordert die Komplexität der Berechnung leicht zu handhabende Verteilungen für die Modellierung, da sonst die Laufzeit nicht für die Praxis anwendbar wäre. Daher wird von komplexeren Verteilungen für die Modellierung abgesehen.

Alternative Verteilung zur Modellierung der Öffnungszahlen

Für die Verteilung der Öffnungszahlen wurde untersucht, ob eine Negative Binomialverteilung besser geeignet ist als die verwendete Poissonverteilung. Es ist bekannt, dass die Summe negativ binomialverteilter Zufallsvariablen wieder einer negativen Binomialverteilung folgen. Die Poissonverteilung ist ebenfalls reproduktiv. Zur Auswahl der geeigneten Verteilung wurde daher die empirische Verteilung der Öffnungszahlen der Kunden pro Monat und Kunde betrachtet. Anschließend wurden die Parameter der Verteilung mit der Summe der quadratischen Abweichungen der Wahrscheinlichkeitsfunktionen geschätzt und ein Chi-Quadrat-Test zur Bestimmung der Güte der Anpassungen durchgeführt. Sowohl durch die grafische Darstellung in Abbildung 12 als auch durch einen höheren Chi-Quadrat-Wert von 922.97 für die NB-Verteilung gegenüber einem Chi-Quadrat-Wert von 829.72 für die Poissonverteilung, wurde die Poissonverteilung als geeignetere Variante für die Modellierung ausgewählt. Es ist auffallend, dass beide Verteilungen einen relativ hohen Chi-Quadrat-Wert aufweisen. Da komplexere Verteilungen aufgrund des hohen Rechenaufwands nicht anwendbar sind, wird dennoch die Poissonverteilung als geeignetste Verteilung für die Modellierung bestimmt.

Tests mit Vergleichskunden

Des Weiteren wurden zur Validierung der Ergebnisse der Parameterschätzungen Tests mit Vergleichskunden durchgeführt. Die Kunden wurden so ausgewählt, dass sie im November 2021 ihren ersten Kauf getätigt haben und nach 6 Monaten mindestens einen weiteren. Nach der Auswahl der Daten wie in 6.1 beschrieben, stehen noch Daten von 552 Kunden für die Validierung zur Verfügung.

Zum Test der Modellspezifikationen wird getestet, ob eine Schätzung der Modellparameter für die Testkunden ebenfalls gute Einblicke in die Kundenbindung geben kann, oder ob beispielsweise eine Festlegung der Poissonverteilung für die Modellierung der Öffnungen eine gute Anpassung des Modells nicht zulassen würde.

Zunächst werden die einzelnen Parameterschätzungen für die Testkunden mit den Parameterschätzungen der vorher verwendeten Kunden diskutiert. Die Parameterschätzungen zum Öffnungsverhalten der Testkunden sind sehr ähnlich zu den Schätzungen mit den in dieser Arbeit verwendeten Kunden. Ebenfalls sind die Einteilungen der verschiedenen Kundengruppen und die Parameterschätzungen zu dem Kaufverhalten sehr ähnlich. Bei der Berechnung der absoluten logarithmierten Likelihood-Berechnung muss beachtet werden, dass eine geringere Menge an Datenpunkten bei den Testkunden vorhanden ist. Nach Skalierung bezüglich der Datenpunkte kann festgestellt werden, dass die absolute logarithmierte Wahrscheinlichkeit um 0.2% höher ausfällt. Aufgrund dieses geringen Unterschieds in der Modellgüte, wird gefolgert, dass die Modellspezifikationen auch für die Modellierung der Kundenbindung anderer Kunden des Thalia Theaters verwendet werden können.

Des Weiteren wird getestet, ob die in dieser Arbeit festgelegten Parameter einen ähnlichen absoluten logarithmierten Likelihood-Wert wie für die Ausgangsdaten annimmt. Mit den verwendeten Parametern wurde ein Likelihood-Wert von 15274.15 erzielt. Das Modell kann durch Anpassung der Parameter bezüglich der Testkunden einen Likelihood-Wert von 14632.94 erzielen. Dies deutet darauf hin, dass für eine bessere Schätzung der Parameter mehr Kunden und längere Betrachtungszeiträume verwendet werden müssen.

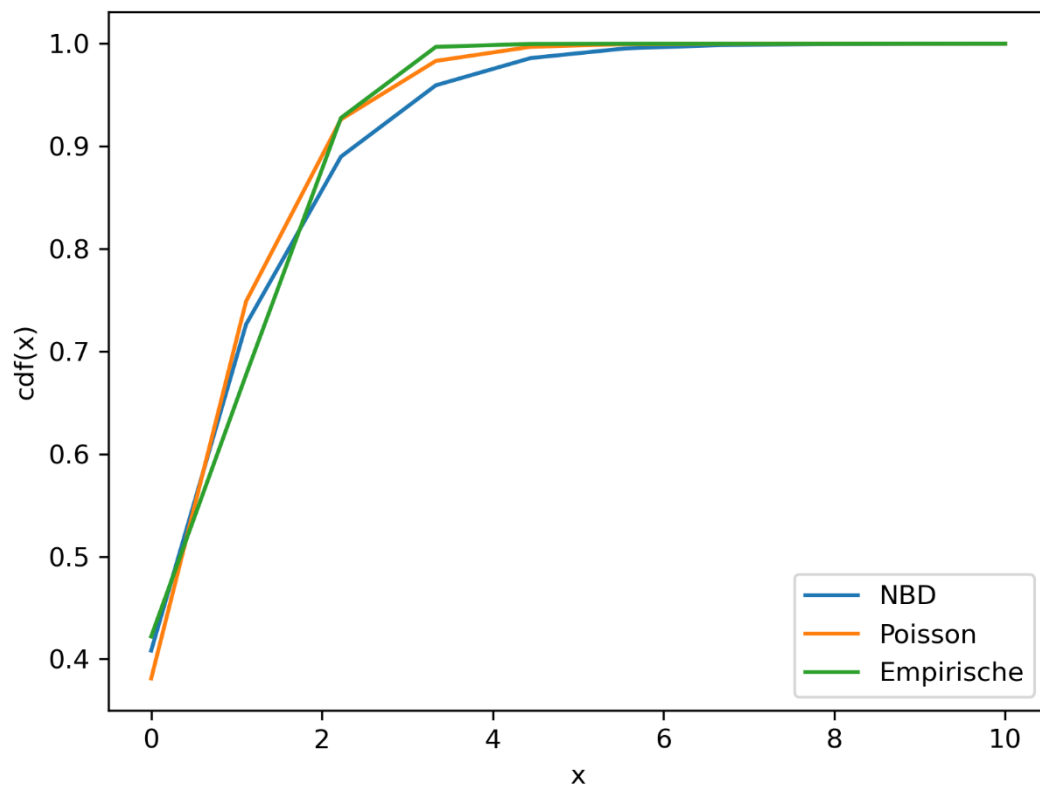


Abbildung 12: Vergleich der Poissonverteilung, NBD und der empirischen Verteilung der Öffnungszahlen

7 Schlussfolgerungen, Beschränkungen und zukünftige Forschung

E-Mail-Marketing-Programme werden in verschiedenen Branchen intensiv genutzt, um mit Kunden in Kontakt zu treten. Die gängige Praxis in der Branche, um die Wirksamkeit eines E-Mail-Marketing-Programms zu messen, besteht darin, die Kundenreaktion auf E-Mails wie die Öffnungsrate zu untersuchen. Es wurde jedoch gezeigt, dass es irreführend sein kann, nur die Öffnungsrate zu berücksichtigen. Diese empirische Arbeit zeigt, dass einige derjenigen, die E-Mails besonders häufig öffnen, in ihren Kaufverhaltensweisen am inaktivsten sind. Wenn Unternehmen ihre Ressourcen ausschließlich auf Grundlage der E-Mail-Öffnungsrate zuweisen, könnten sie möglicherweise eine Gruppe von Kunden übersehen, die zwar nicht auf E-Mails reagieren, aber dennoch relativ aktiv im Kaufverhalten sind.

Das Ziel dieser Arbeit besteht nicht darin, die Aufmerksamkeit der Unternehmen von der E-Mail-Öffnungsrate abzulenken. Tatsächlich zeigt diese Arbeit, dass es im Durchschnitt eine positive Korrelation zwischen dem Öffnen von E-Mails und dem Kaufverhalten gibt. Es wird jedoch empfohlen, dass Unternehmen neben den E-Mail-Reaktionsraten auch das Kaufverhalten der Kunden betrachten sollten. Wenn das Ziel darin besteht, die langfristige Rentabilität zu maximieren, sollten Unternehmen über das optimale Maß an E-Mail-Kontakten informiert sein, die sie mit ihren Kunden herstellen sollten.

Eine der Einschränkungen dieser Studie besteht darin, dass der Inhalt der E-Mails nicht beobachtet wird. Daher wird sich nicht auf die E-Mail-Strategie in Verbindung mit Bedürfnissen der Kunden, wie beispielsweise Cross-Selling und Up-Selling, konzentriert. Wenn Daten zum E-Mail-Inhalt verfügbar sind, könnte zukünftige Forschung nicht nur die Auswirkungen des E-Mail-Inhalts auf die Reaktion der Kunden und das Kaufverhalten berücksichtigen, sondern auch Anleitung für das Unternehmen geben, wie es das E-Mail-Marketing basierend auf einer personalisierten Anpassung des E-Mail-Inhalts gezielter gestalten kann (vgl. Zhang et al. 2017, S. 865).

Eine weitere Einschränkung stellt der relativ kurze Beobachtungszeitraum für die verwendeten Daten dar. Aufgrund der Corona Beschränkungen war ein Betrieb des Thalia Theaters bis zum Oktober 2021 stark eingeschränkt. Dadurch dass im Modell, Variablen verwendet werden, die den Einfluss der Zeit zwischen Käufen und der Zeit zwischen Öffnungen modellieren, wäre eine längerer Beobachtungszeitraum wichtig für eine bessere Schätzung des Einflusses dieser Variablen gewesen.

Zudem stellt die Rechenkapazität des für die Berechnungen verwendeten Laptops eine große Herausforderung dar. Dadurch, dass das Modell eine Vielzahl von Parametern mit komplexen

Zusammenhängen beinhaltet, stellt die Anwendung des Modells auch Anforderungen an die technischen Mittel.

Zukünftige Forschung kann das Modell zur Analyse des Verhaltens von Neukunden oder Kunden, die seit langer Zeit keinen Kauf mehr tätigten, beinhalten. Gerade die Interaktion mit Willkommensmails kann Aufschlüsse darüber geben, wie eine grundsätzliche Einstellung von Neukunden gegenüber dem Unternehmen aussehen.

Literaturverzeichnis

Bruhn, Manfred (Hg.) (2005): Handbuch Kundenbindungsmanagement. Strategien und Instrumente für ein erfolgreiches CRM. 5., überarb. und erw. Aufl. Wiesbaden: Gabler.

Franziska Pinkert; Pascal Thomann; Stephanie Frömel (2022): Impulse Webinar. Integration von Ticketing- und CRM-System. Projekterfahrungen aus dem Theater Basel. kulturplaner; Theater Basel. kulturplaner, 23.11.2022. Online verfügbar unter <https://www.youtube.com/watch?v=uH3czEOUOVU>, zuletzt geprüft am 19.06.2023.

Freie und Hansestadt Hamburg (2021): Haushaltsplan 2021/2022. Einzelplan 3.3, Behörde für Kultur und Medien. Online verfügbar unter <https://www.hamburg.de/fb/haushaltsplaene/14735932/doppelhaushalt-2021-2022/>, zuletzt geprüft am 17.06.2023.

Freie und Hansestadt Hamburg (2022): Haushaltsplan 2023/2024.

hamburg.de (2023): Thalia Theater Hamburg - Öffnungszeiten, Bilder, Informationen. Online verfügbar unter <https://www.hamburg.de/sehenswuerdigkeiten/1093226/thalia-theater/>, zuletzt aktualisiert am 17.06.2023, zuletzt geprüft am 17.06.2023.

Inxmail GmbH (2023): E-Mail-Marketing-Benchmark 2023.

Kornfeld, Michael (2023): E-Mail Blog: Die Öffnungs-Erkennung: Hintergrund & Interpretation. Michael Kornfeld. Online verfügbar unter https://www.dialog-mail.com/email-blog/2021/20210509_newsletter-oeffnungs-erkennung.php, zuletzt aktualisiert am 20.06.2023, zuletzt geprüft am 20.06.2023.

Netzer, Oded; Lattin, James M.; Srinivasan, V. (2008): A Hidden Markov Model of Customer Relationship Dynamics. In: *Marketing Science* 27 (2), S. 185–204. DOI: 10.1287/mksc.1070.0294.

Rabiner, L.; Juang, B. (1986): An introduction to hidden Markov models. In: *IEEE ASSP Mag.* 3 (1), S. 4–16. DOI: 10.1109/MASSP.1986.1165342.

Statista (2023): Markt für darstellende Künste: Anzahl der Unternehmen | Statista. Online verfügbar unter <https://de.statista.com/statistik/daten/studie/165749/umfrage/anzahl-der-unternehmen-im-markt-fuer-darstellende-kuenste-seit-2003/>, zuletzt aktualisiert am 18.06.2023, zuletzt geprüft am 18.06.2023.

Stewart, Brandon (2021): Wird Apples Mail Privacy Protection das E-Mail-Marketing für immer verändern? | Emarsys. Online verfügbar unter <https://emarsys.com/de/learn/blog/wird-apples-mail-privacy-protection-das-e-mail-marketing-fur-immer-verandern/>, zuletzt aktualisiert am 12.07.2021, zuletzt geprüft am 20.06.2023.

swissfuture (2014): Die Zukunft der Kulturszene. In: *Magazin für Zukunftsmonitoring* 41 (3). Online verfügbar unter https://kulturmanagement.philhist.unibas.ch/fileadmin/user_upload/kulturmanagement/Dokumente_-_Downloads/Swissfuture_3_14_Kultur.pdf.

Zhang, Xi; Kumar, V.; Cosguner, Koray (2017): Dynamically Managing a Profitable Email Marketing Program. In: *Journal of Marketing Research* 54 (6), S. 851–866. DOI: 10.1509/jmr.16.0210.

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit mit dem Titel

Data Analytics in der Tourismusbranche am Bei-spiel der Firma Wilken

selbständig angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Ich bin mir bewusst, dass eine unwahre Erklärung rechtliche Folgen haben wird.

Ort, Datum

Vorname Nachname