**exp**

**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

Pof. Dr. Mitja Stadje

# **Sperrvermerk**

(nur bei Arbeiten in Kooperation mit Unternehmen)

Diese Arbeit enthält Informationen und Daten, welche die Belange und Interessen der *XYZ GmbH* 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

Abkürzungsverzeichnis VII

1 Einleitung 1

2 Problemstellung 3

3 Theoretische Grundlagen 4

3.1 Kundenbindung im Kulturbereich 5

3.2 Das Thalia Theater 9

3.3 Email-Marketing des Thalia Theaters im Zeitraum 11

3.4 Definition eines Hidden-Markov-Models (HMM) 13

3.4.1 Intuition 13

3.4.2 Elemente eines Hidden-Markov-Models 16

3.5 Öffnungserkennung 18

3.5.1 Erkennung einer Öffnung 18

3.5.2 Rolle der E-Marketing Suite der Firma Wilken 18

3.6 Datensammlung 19

3.7 Modellierungs-Framework 20

3.8 Überblick über das Modell 21

3.9 Startverteilung 21

3.10 Übergangsmatrix 22

3.11 Bedingte Email-Öffungsanzahl (CEOM) 23

3.12 Modell der Kaufanzahl (CPM) 24

3.13 Die Korrelation zwischen Kaufetscheidung und Email Öffnungen 25

3.14 Berichtigung der Endogenität 27

3.15 Model estimation 27

4 Verwandte Arbeiten 29

5 Evaluation 30

5.1 Datenauswahl 30

5.2 Modellidentifikation 31

5.3 Parameterbestimmungen 34

5.4 Bestimmung des Kundenzustands 41

5.5 Zusammenfassung 43

5.6 Robustheitstests 43

6 Optimales E-Mailmarketing 47

7 Schlussfolgerungen, Beschränkungen und zukünftige Forschung 52

Anlage VII

Literaturverzeichnis VIII

Ehrenwörtliche Erklärung IX

# Abbildungsverzeichnis

[Abbildung 1: Konzeptualiserung der Kundenbindung (Bruhn 2005), S. 8 6](#_Toc138955080)

[Abbildung 2: Anzahl der Unternehmen im Markt für darstellende Künste in Deutschland von 2003 bis 2021 (Statista 2023) 8](#_Toc138955081)

[Abbildung 3: Ticketverkäufe des Thalia Theaters in den Jahren 2018 bis 2022 nach Quartalen 10](#_Toc138955082)

[Abbildung 4: Aggregierte Verkäufe pro Monat (2014 bis 2019) 11](#_Toc138955083)

[Abbildung 5: Versendete Mails im Zeitraum Oktober 2021 bis Dezember 2022 12](#_Toc138955084)

[Abbildung 6: Einbindung der E-Marketing Suite in die CRM-Landschaft des Theater Basel (Franziska Pinkert et al. 2022) 19](#_Toc138955085)

[Abbildung 7: Histogramm derTicketverkäufe (Januar 2021 bis Dezember 2022) 30](#_Toc138955086)

[Abbildung 8: Histogramm der Käufe pro Monat und Kunde (im Zeitraum Oktober 2021 bis Dezember 2022) 32](#_Toc138955087)

[Abbildung 9: Vergleich der Verteilungsfunktionen (CDF) einer ZINBD-, NBD- und Poissonverteilten ZV mit geschätzten Parameterwerten gegenüber der empirischen Verteilungsfunktion 33](#_Toc138955088)

[Abbildung 10: Wahrscheinlichkeitsfunktion der Käufe für verschiedene Werte von lambda 37](#_Toc138955089)

[Abbildung 11: Wahrscheinlichkeiten von zufälligen Kunden in Zustand 1 zu sein 43](#_Toc138955090)

[Abbildung 12: Vergleich der Poissonverteilung, NBD und der empirischen Verteilun der Öffungszahlen 46](#_Toc138955091)

# Tabellenverzeichnis

[Tabelle 1: Startparameter und Grenzen (für die einzelnen Werte in den Parametern) 34](#_Toc138955067)

[Tabelle 2: Parameterschätzungen (noch nicht aktualisiert) 36](#_Toc138955068)

[Tabelle 3: Relevante Parameter für die Modellierung des Kaufverhaltens 36](#_Toc138955069)

[Tabelle 4: Einfluss der E-Mailkampagnen auf das Kaufverhalten 38](#_Toc138955070)

[Tabelle 5: Relevante Parameter für die Modellierung der E-Mailöffnungen 38](#_Toc138955071)

[Tabelle 6: Veranschaulichung der Übergangsmatrix anhand eines „typischen“ Kunden (eine Öffnung im letzten Monat und kein Kauf) 39](#_Toc138955072)

[Tabelle 7: Veranschaulichung der Übergangsmatrix anhand eines Kunden mit einer Öffnung im letzten Monat und einem Kauf 39](#_Toc138955073)

[Tabelle 8: Veranschaulichung der Übergangsmatrix anhand eines Kunden mit keiner Öffnung im letzten Monat und keinem Kauf 40](#_Toc138955074)

# 

# Abkürzungsverzeichnis

Hinweis:

In das Abkürzungsverzeichnis werden alle Abkürzungen aufgenommen, die im Text verwendet werden und die gleichzeitig nicht im Duden als Abkürzungen aufgeführt sind. Deshalb werden beispielsweise „Aufl.“ und „GmbH“ nicht in das Verzeichnis aufgenommen. Die Formatierung ist dieselbe wie im Abbildungs- und Tabellenverzeichnis.

Außerdem sollten Abkürzungen aus dem Abkürzungsverzeichnis einmal eingeführt (z. B. Social Network Analysis (SNA)) und anschließend der zugehörige Begriff nur noch abgekürzt verwendet werden.

|  |  |
| --- | --- |
| HMM | Hidden Markov Model |
| Bfai | Bundesagentur für Außenwirtschaft |
| KI | Künstliche Intelligenz |
| überarb. | überarbeitete |
| SNA | Social Network Analysis |

# 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 zugrundelegende 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 Unterehmen 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 Apassung 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.

# Problemkontext

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 Staastheater ist es finanziell abhängig von dem Haushalt der Stadt Hamburg. Dadurch dass ein festgesetztes Budget für den Betrieb des Unterehmens vorgegeben wird, spielt insbesondere eine gute Kostenpolitik eine wichtige Rolle bei wirtschaftlichen Erwägungen. Somit ist eine richtige Allokation von Marketingresourcen sehr wichtig für das Unternehmen. Zudem steht das Thalia Theater wie andere Anbieter für die Freizeitgestaltung vor der Herausforderung eines immer größer werdenden Agebots 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 Marketingresourcen 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.

# Theoretische Grundlagen

## 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 werden die strategische Bedeutung und der Begriff der Kundenbindung erläutert. Danach wird die Wirkungskette der Kundebindung 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.“

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 Abbildung 1 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 des 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 mi 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).

Abbildung 2 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 Einschräkungen wärend der Corona Pandemie in den Jahren 2020 und 2021 stellten große Herausforderungen für den Markt dar und betont die Wichtigkeit einer nachhaltigen Wirtschaftsweise für die Betriebe.

Ein Bild, das Text, Reihe, Diagramm, Quittung enthält.

Automatisch generierte Beschreibung

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

**Wirkugskette der Kundenbindung**

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

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 zur 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 Email-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 wie 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.

Ein Bild, das Text, Screenshot, parallel, Zahl enthält.

Automatisch generierte Beschreibung

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

## 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 wude 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 ist 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).

**Wirtschaftlichkeit**

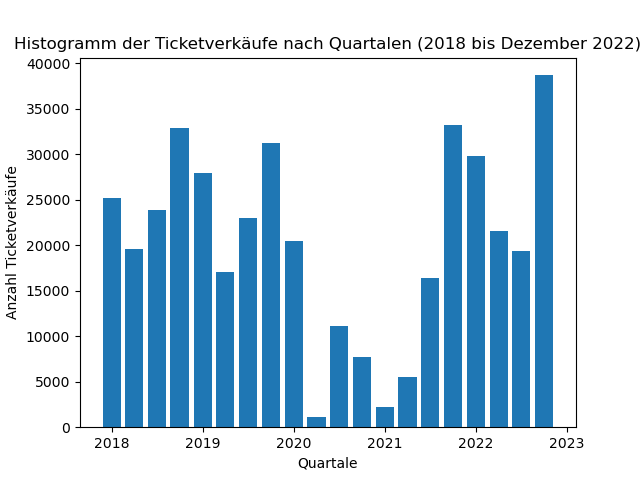
Der Kostendeckungsgrad des Thalia Theaters betrug im Jahr 2018 22,5% und stieg im Jahr 2019 leicht auf 22,8% (Freie und Hansestadt Hamburg 2021, S. 45). Dies zeigt deutlich, dass das Thalia Theater für den Erhalt des Betriebs stark auf die Förderungen der Stadt Hamurg angewiesen ist. Die Behörde für 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).

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

**Saisonaler Effekt**

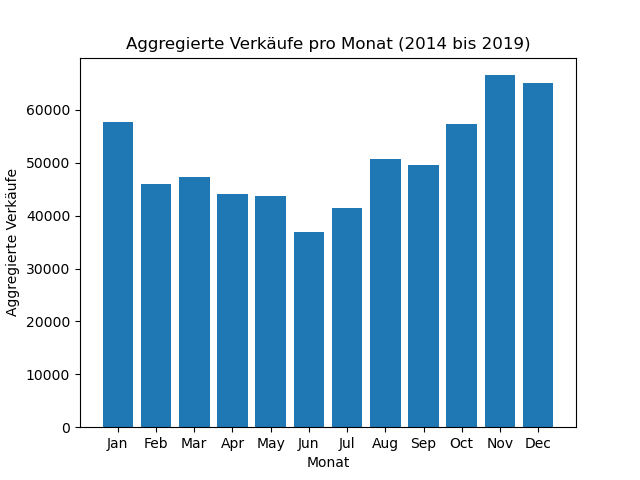


Abbildung : Aggregierte Verkäufe pro Monat (2014 bis 2019)

**Wirtschaftlicher Erfolg nach Corona**

## Email-Marketing des Thalia Theaters im Zeitraum

Das folgende Unterkapitel untersucht das Emailmarketing des Thalia Theaters im Zeitraum von Oktober 2021 bis Dezember 2022. Dabei liegt der Fokus auf zwei zentralen Aspekten: der Häufigkeit des Emailversands und der Gestaltung der Emails. Durch die Analyse dieser Themen werden Einblicke in die Emailkommunikation des Theaters gewonnen und deren Auswirkungen auf die Empfänger untersucht.

**Häufigkeit des Emailversands**

Im Zeitraum von Oktober 2021 bis Dezember 2022 wurden insgesamt 30 Kampagnen vom Thalia Theater durchgeführt, um das Emailmarketing zu gestalten und die Zielgruppe effektiv anzusprechen. Abbildung 5 visualisert 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 Kampgnen im März 2022 als eine zusätzliche Kampagne zur „Solidarität der Ukraine“ geschaltet wurde.

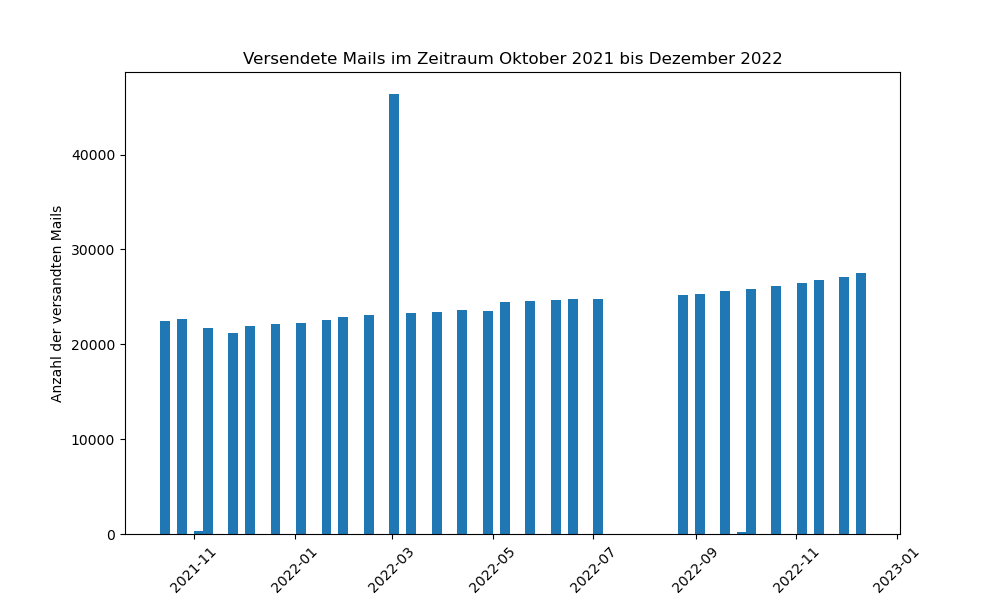


Abbildung : Versendete Mails im Zeitraum Oktober 2021 bis Dezember 2022

**Gestaltung der E-mailkampagnen**

Die E-mailkampagnen 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 duch 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.

## Definition eines Hidden-Markov-Models (HMM)

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 dieser Masterarbeit werden wir die Anwendung von HMMs im Marketing genauer untersuchen und deren Potenzial zur Steigerung des Marketingerfolgs erforschen.

### Intuition

(Rabiner und Juang 1986, S. 5–7)

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1165342

Ein Hidden Markov Model (HMM) ist ein zweifacher stochastischer Prozess mit einem zugrunde liegenden stochastischen Prozess, der nicht beobachtbar ist (er ist verborgen), sondern 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.

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 eine Münzwurf- (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 man beobachtet nur die Ergebnisse der Münzwürfe, 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 1a 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 Münzwurfergebnissen wird in Abbildung 1b 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 1a 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 auf lange Sicht (z. B. durchschnittliche Anzahl von "Kopf" oder "Zahl") der Beobachtungssequenzen des HMMs in Abbildung 1c dieselben sind wie die der Modelle in den Abbildungen 1a und 1b. 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 1c. 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.

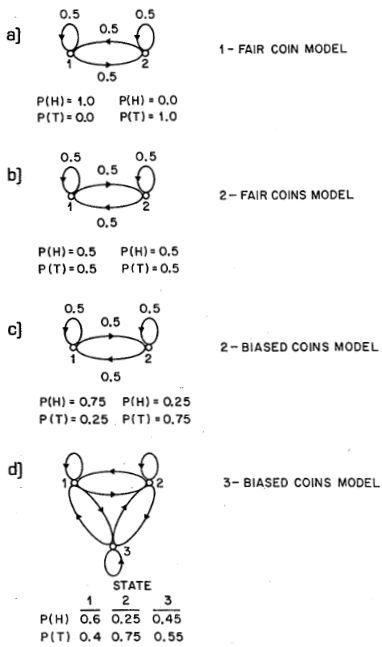


Abbildung 1: Modellierung von verborgenen Münzwurf-Experimenten (Rabiner und Juang 1986, S. 6)

Das Modell in Abbildung 1d, 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 1d 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 angehalten 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.

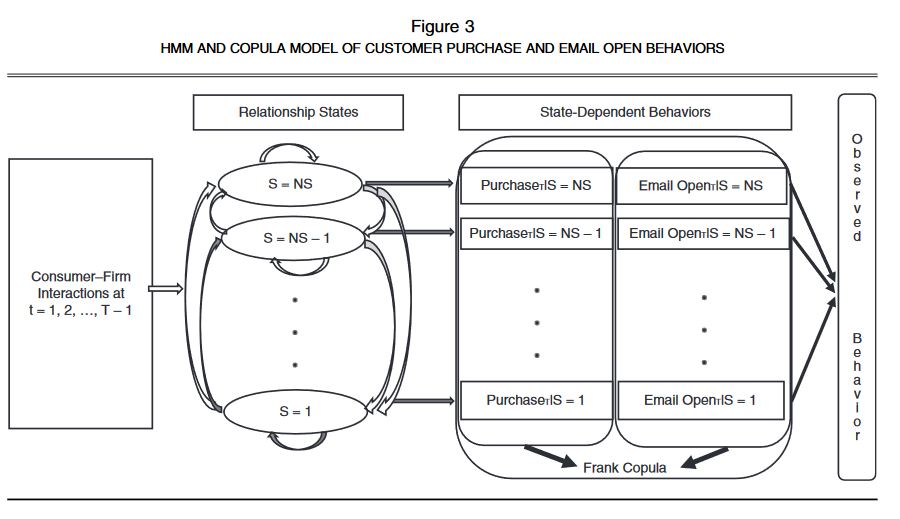
Ein letzter Punkt betrifft die Größe der Beobachtungssequenz. Wenn wir auf eine kleine, endliche Beobachtungssequenz beschränkt sind, können wir möglicherweise die optimalen Modellparameter nicht zuverlässig schätzen. (Denken Sie an den Fall, dass tatsächlich 10 Münzen verwendet werden, aber nur 50-100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

### Elemente eines Hidden-Markov-Models

(Rabiner und Juang 1986, S. 7)

Wir erklären nun die Elemente und den Mechanismus des HMMs, das wir in dieser Arbeit behandeln:

1. Es gibt eine endliche Anzahl von Zuständen in dem Modell, wir bezeichnen diese als S = 1, …, NS (für „state“, dt. „Zustand). Jeder Zustand spiegelt ein anderes Niveau der Kundenbindung wieder.
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 Unternemen 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.



## Öffnungserkennung

### Erkennung einer Öffnung

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 Vorschaufenster 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).

### Rolle 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-Mailkampagnen bei Kulturunternehmen einnimmt. Die E-Marketing Suite bietet Schnittstellen zu anderen Anbietern im CRM-Markt an, wie in Abbildung 6 dargestellt. So erfasst das Ticketsystem der Firma Secutix während des Ticketing-Prozesses die Transaktionsdaten und übermittelt diese an die Firma kulturplanner. kulturplanner 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. 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-Mailkampagnen ebenfalls über die E-Marketing Suite verläuft, enthalten die Kundendaten in der E-Marketing Suite ebenfalls Informationen zu den Öffnungen, Klicks und Informationen, wann und ob eine E-Mail erhalten wurde.

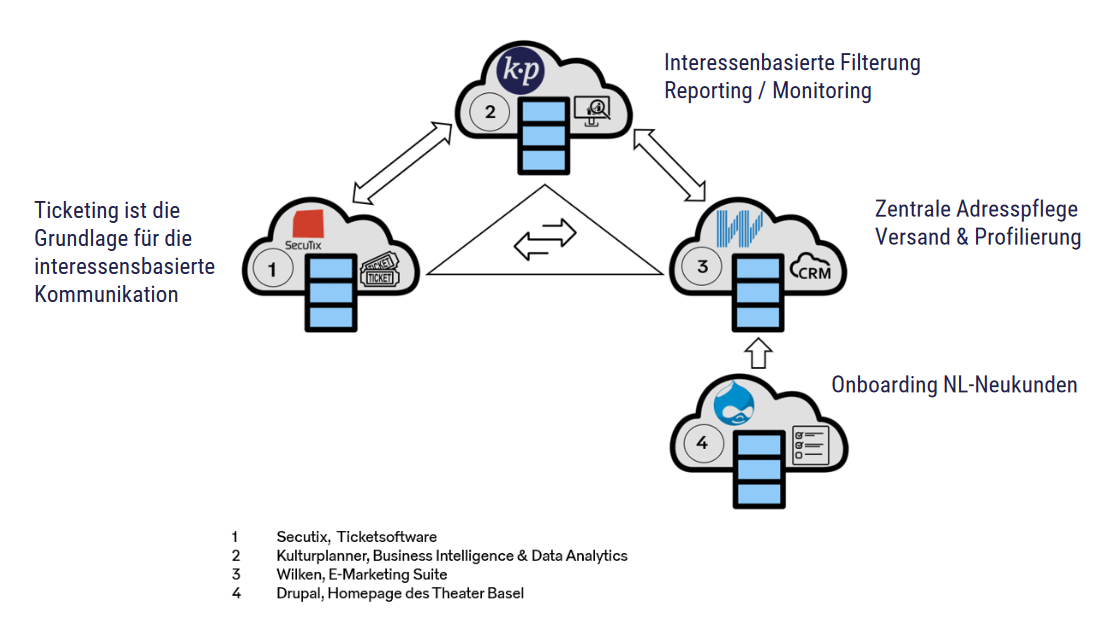


Abbildung : Einbindung der E-Marketing Suite in die CRM-Landschaft des Theater Basel (Franziska Pinkert et al. 2022)

## Datensammlung

## Modellierungs-Framework

file:///C:/Users/felix.oechslein/Desktop/MA/An\_introduction\_to\_hidden\_Markov\_models-1.pdf

Angesichts der Form des HMM, die in dem vorherigen Abschnitt diskutiert wurde, gibt es drei wesentliche Probleme, die gelöst werden müssen, damit das Modell in realen Anwendungen nützlich ist. Diese Probleme sind wie folgt:

Problem 1: Gegeben sei die Beobachtungssequenz O = O1, O2, ..., O7 und das Modell A = (A, B, π). Wie berechnen wir Pr(O|A), die Wahrscheinlichkeit der Beobachtungssequenz unter dem Modell?

Problem 2: Gegeben sei die Beobachtungssequenz O = O1, O2, ..., Or. Wie wählen wir eine Zustandssequenz | = i1, i2, ..., i7, die in irgendeinem sinnvollen Sinne optimal ist?

Problem 3: Wie passen wir die Modellparameter (A, B, π) an, um Pr(O|A) zu maximieren?

Problem 1 ist das Bewertungsproblem: Gegeben ein Modell und eine Sequenz von Beobachtungen, wie können wir die Wahrscheinlichkeit berechnen, dass die beobachtete Sequenz vom Modell erzeugt wurde? Wir können das Problem auch so betrachten: Gegeben ein Modell und eine Sequenz von Beobachtungen, wie können wir das Modell "bewerten" oder "benoten"? Diese Sichtweise ist sehr nützlich. Wenn wir den Fall betrachten, in dem wir mehrere konkurrierende Modelle haben (z. B. die vier Modelle in Abbildung 1 für das Münzwurfexperiment), ermöglicht uns die Lösung von Problem 1 die Auswahl des Modells, das am besten mit den Beobachtungen übereinstimmt.

Problem 2 besteht darin, den verborgenen Teil des Modells, nämlich die Zustandssequenz, zu ermitteln. Dies ist ein typisches Schätzungsproblem. Wir verwenden in der Regel ein Optimalitätskriterium, um dieses Problem bestmöglich zu lösen. Leider gibt es, wie wir sehen werden, mehrere mögliche Optimalitätskriterien, die angewendet werden können, und daher hängt die Wahl des Kriteriums stark von der beabsichtigten Verwendung der ermittelten Zustandssequenz ab. Eine typische Verwendung der wiederhergestellten Zustandssequenz besteht darin, Informationen über die Struktur des Modells zu gewinnen und Durchschnittsstatistiken, Verhalten usw. innerhalb einzelner Zustände zu erhalten.

Problem 3 besteht darin, die Modellparameter so zu optimieren, dass sie am besten beschreiben, wie die beobachtete Sequenz entsteht. In diesem Fall nennen wir es eine Trainingsssequenz, da sie dazu verwendet wird, das Modell zu trainieren. Das Training ist das entscheidende Problem für die meisten Anwendungen von HMM's, da es uns ermöglicht, die Modellparameter optimal an beobachtete Trainingsdaten anzupassen, d.h. die besten Modelle für reale Phänomene zu erstellen.

Im Einklang mit früheren Untersuchungen verwenden wir ein HMM, um die um die Zustände der Kunden-Firmen-Beziehung und die Übergänge zwischen diesen Zuständen. Ein HMM beschreibt einen Markov-Prozess mit diskreten latenten Zustände. Ein HMM ist ein stochastisches Modell, das zur Erfassung der Übergänge zwischen diesen latenten Zuständen zu erfassen und diese Zustände in beobachtete Verhaltensweisen. In der Marketingliteratur werden HMMs häufig verwendet, um Beziehungen zwischen Kunden und Unternehmen verwendet (z. B. Kumar et al. 2011; Luo und Kumar 2013; Montoya, Netzer, und Jedidi 2010; Netzer, Lattin, und Srinivasan 2008). Im Kontext des E-Mail-Marketings, verwenden wir ein HMM, um zwei beobachtete Kundenverhaltensweisen zu untersuchen zu untersuchen: Käufe und E-Mail-Öffnungen.4 In unserer Umgebung entsprechen die latenten Zustände des Markov-Prozesses verschiedenen Ebenen von Kunden-Firmen-Beziehungen, die eine unterschiedliche Kauf- und E-Mail-Öffnungsaktivität der Kunden führen. Da es außerdem eine mögliche Korrelation zwischen dem Kauf- und dem E-Mail-Öffnungsverhalten der Kunden besteht Kauf- und E-Mail-Öffnungsverhalten besteht, erfassen wir die Korrelation zwischen den beiden Verhaltensweisen durch eine Frank-Copula-Funktion (für die grafische graphische Darstellung des von uns vorgeschlagenen HMM- und Copula-Modells des und des E-Mail Öffnungsverhaltens, siehe Abbildung 3). In diesem Abschnitt erörtern wir zunächst die Grundzüge unseres HMM. Zweitens spezifizieren wir unsere bedingten (auf latente Zustände des Markov-Prozesses) E-Mail-Öffnungs- und Kaufmodelle. Drittens veranschaulichen wir, wie wir die Korelation zwischen zustandsabhängigem abhängigen E-Mail-Öffnungs- und Kaufverhalten durch unsere Frank-Kopula-Komponente erfassen. Viertens diskutieren wir die Modellschätzung. Abschließend diskutieren wir die Modellidentifikation.

## Überblick über das Modell

Sei Oit die Anzahl der E-Mails, die Kunde i zur Zeit t öffnet, und Yit die Anzahl der Einkäufe, die Kunde i zur Zeit t tätigt. Für Kunden i modellieren wir die Abfolge der Beobachtungen

[(Yi1 = yi1, Oi1 = oi1), ..., (Yit = yit, Oit = oit)] unter Verwendung eines HMM, das durch (1) die Anfangszustandsverteilung (ψi), (2) eine Sequenz von Übergangswahrscheinlichkeiten (Qit) und (3) einen Vektor von Wahrscheinlichkeiten, die die latenten Zustände mit den beobachteten Kauf- und E-Mail-Öffnungsverhalten in Beziehung setzen (Hit), charakterisiert wird.

## Startverteilung

Zu einem beliebigen Monat t sei s der Grad der Beziehung des Kunden i. Sei ψis die Wahrscheinlichkeit, dass Kunde i zu Beginn in Beziehungszustand s ist, wobei ψis ≥ 0 und , wobei NS die Anzahl der latenten Markov-Zustände ist. In dieser Studie gehen wir davon aus, dass alle Kunden im ersten Monat mit dem niedrigsten Beziehungszustand beginnen. Daher ist ψis´ = (ψi1, … , ψi,NS ) = (1, 0, ..., 0).

## Übergangsmatrix

In unserem vorgeschlagenen HMM-Framework erlauben wir Kunden, in jeden Beziehungszustand s = 1, 2, ..., NS zu wechseln. Wie bei Kumar et al. (2011) verwenden wir eine multinomiale Logit-Spezifikation, um diesen Übergangsprozess zu formulieren. Wir definieren die Übergangsmatrix wie folgt:

Ein Bild, das Text, Tisch enthält.

Automatisch generierte Beschreibung

Hierbei ist qit,s→s0 die Wahrscheinlichkeit, dass Kunde i vom Zustand s zur Zeit t-1 in den Zustand s´ zur Zeit t wechselt, wobei 0 ≤ qit,s→s´ ≤ 1 für alle s,s´ gilt und 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 (Uit,s→s0) wie folgt:

|  |  |
| --- | --- |
|  | () |

Hierbei bezeichnen eit,s→s’ und Vit,s→s’ die stochastischen und deterministischen Komponenten des indirekten Nutzens. Wir nehmen an, dass eit,s→s’ für i = 1, ..., N; t = 1, ..., T; s, s’ = 1, ..., NS i.i.d. Gumbel-verteilt mit Lageparameter 0 und Skalenparameter 1 sind. Wir operationalisieren die deterministische Komponente vit,s→s’ wie folgt:

|  |  |
| --- | --- |
|  | () |

wobei αs→s’ der intrinsische Nutzen ist, der mit der Transition von Zustand s zu s’ zum Zeitpunkt t einhergeht. Xi,t−1 enthält folgende Variablen: I[Oi,t−1 > 0], I[Yi,t−1 > 0], EMi,t−1, EM i,t-12 , wobei I[A] die Indikatorfunktion ist, die den Wert 1 annimmt, wenn das Ereignis A eintritt, und sonst den Wert 0 annimmt. Oi,t−1 ist die verzögerte Anzahl geöffneter E-Mails, Yi,t−1 ist die verzögerte Anzahl von Käufen und EMi,t−1 ist die Anzahl der vom Unternehmen zum Zeitpunkt t−1 gesendeten E-Mails. γs→s’ = [γ1,s→s’, γ2,s→s’, γ3,s→s’, γ4,s→s0] bezeichnet den Vektor der entsprechenden Antwortkoeffizienten. Wir normalisieren den deterministischen Nutzen für Kunden i zum Zeitpunkt t, um zur niedrigsten Stufe (Vit,s→1) den Wert 0 für Identifikationszwecke zu erhalten (d. h., Uit,s→1 = eit,s→1, wobei eit,s→1 für i = 1, ..., N; t = 1, ..., T; s = 1, ..., 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 (qit,s→s0), zur folgenden bekannten Multinomial-Logit-Share-Funktion:

|  |  |
| --- | --- |
|  | () |

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

## Bedingte Email-Öffungsanzahl (CEOM)

Wir nehmen an, dass die Anzahl der E-Mails, die Kunde i (im Zustand s) zum Zeitpunkt t öffnet, einer Binomialverteilung mit Parametern EMit und pit|s folgt, gegeben durch:

|  |  |
| --- | --- |
|  | () |

wobei pit|s die bedingte Wahrscheinlichkeit ist, dass Kunde i (im Zustand s) im Monat t eine E-Mail öffnet. Wir modellieren pit|s als Funktion des vergangenen E-Mail-Öffnungsverhaltens der Kunden wie folgt:

|  |  |
| --- | --- |
|  | () |

wobei αio|s und βio 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 dem letzten geöffneten E-Mail (d. h. LOit). Wir verwenden den natürlichen Logarithmus von LOit, um den abnehmenden Effekt zu erfassen. Zur Identifikation legen wir die folgenden Beschränkungen fest: (1) αio|s+1 = αio|s + exp(∇αio|s+1), wobei ∇αio|s+1 ein Parameter ist, der aus den Daten geschätzt wird, und (2) βio 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 αio|s als auch βio kundenspezifisch zu sein, um die unbeobachtete Kundenheterogenität zu kontrollieren. Wir nehmen an, dass αio|s und βio mit den folgenden Normalverteilungen verteilt sind.

|  |  |
| --- | --- |
| ,  wobei | () |

## 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 φit|s, λit|s und r folgt. Für jede Beobachtung yit 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:

|  |  |
| --- | --- |
|  | () |

wobei r ein Streuparameter ist, für den angenommen wird, nicht von Kovariaten abhängig zu sein. Der bedingte Mittelwert und die Varianz des ZINBD sind gegeben durch E(yit|s) = (1-φit|s)λit|s und Var(yit|s) = (1- φit|s) λit|s (1+ φit|s λit|s + λit|s r), respectively. φit|s und λit|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 (φit|s) als:

|  |  |
| --- | --- |
|  | () |

wobei δ0i|s und δ1i|s den intrinsischen Nutzen des Kaufs und den Effekt der Dauerabhängigkeit im Zustand s wiedergeben, jeweils. Wir erfassen die Dauerabhängigkeit durch die Variable Zeit seit dem letzten Kauf (LYit). Wir verwenden den natürlichen Logarithmus von LYit, um den abnehmenden Effekt zu erfassen. Um die unerfasste Heterogenität zu berücksichtigen, erlauben wir, dass δki|s (für k=0,1) kundenspezifisch sind. Wir nehmen an, dass δki|s (für k=0,1) über Kunden wie folgt normal verteilt sind:

|  |  |
| --- | --- |
|  | () |

wobei ∇δki ~ N(0, σki2) für k = 0,1.

Wir modellieren die zweite Komponente des bedingten Kaufmodells (d. h. bedingter erwartete Anzahl an Käufen [λit|s]) als eine Funktion der Anzahl der E-Mails, die von der Firma gesendet wurden (EMit), gegeben durch

|  |  |
| --- | --- |
|  | () |

Bedingt auf den Zustand s, ist αip|s die intrinsische Neigung zum Kauf und β1,ip|s und β2,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 αip|1 ≤ αip|2 ≤ … ≤ αip|NS und/oder δip|1 ≤ δip|2 ≤ … ≤ δip|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 auf Käufen aktiv sein müssen, lassen wir die Daten zeigen, ob dies wirklich der Fall ist.

Um die unerkannte Heterogenität zu berücksichtigen, erlauben wir αip|s, β1,ip|s und β2,ip|s kundenspezifisch zu sein. Wir gehen davon aus, dass αip|s und βk,ip|s (für k = 1,2) bei den Kunden wie folgt normal verteilt sind:

|  |  |
| --- | --- |
| ,  wobei | () |

## Die Korrelation zwischen Kaufetscheidung und Email 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. Beachten Sie, dass sowohl die Kaufanzahl Yit als auch die E-Mail-Öffnungszahl Oit einer diskreten Verteilung folgen. Es ist nicht einfach, eine bivariate Verteilung zu finden, die die Korrelation zwischen der ZINBD und der BD erfassen kann.

Danaher und Smith (2011) haben als Pioniere den Copula-Ansatz im Marketing eingeführt, um zwei marginale Verteilungen zu verknüpfen, die nicht aus derselben Familie stammen. Copula-Modelle werden in verschiedenen Studien zur Untersuchung multidimensionaler Marketingprobleme verwendet (z. B. Glady, Lemmens und Croux 2015; Kumar, Zhang und Luo 2014; Park und Gupta 2012; Schweidel und Knox 2013; Stephen und Galak 2012). Die meisten Studien konzentrieren sich jedoch darauf, Copula-Modelle zu verwenden, um Variablen aus kontinuierlichen Verteilungsfamilien zu korrelieren. Stephen und Galak (2012) modellieren multivariate Zählvariablen mit einem doppelten Poisson-Modell, konvertieren jedoch zuerst diese diskreten Variablen in kontinuierliche und verwenden die Gauss'sche Copula, um diese konvertierten Variablen zu korrelieren. Im Kontext dieser Studie haben wir es jedoch mit zwei diskreten Variablen zu tun, die durch zwei verschiedene Verteilungen charakterisiert sind: ZINBD und BD. Es ist ungewiss, ob der Ansatz von Stephen und Galak mit solch unterschiedlichen diskreten Verteilungen funktionieren wird. Darüber hinaus ist es aus Effizienzgründen wünschenswert, die Korrelation zwischen den beiden diskreten Variablen direkt zu modellieren, anstatt sie zuerst in kontinuierliche umzuwandeln und sie als kontinuierliche zu behandeln.

Um die Herausforderung der Korrelation diskreter Verteilungen anzugehen, greifen wir auf die Literatur der Mathematik und Statistik zurück. 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 Studie, 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. Es gibt mehrere statistische Anwendungen dieser Idee zur Modellierung von bivariaten Zähldaten. Lee (1999) entwickelt eine bivariate negative BD zur Modellierung von Rugby-League-Punkten unter Verwendung von Frank-Copula. Song (2000) entwickelt ein multivariates Streuungsmodell, das aus der Gaußschen Copula generiert wird. Nikoloulopoulos und Karlis (2010) erfassen die Beziehung zwischen den Kaufzahlen bestimmter Produktkategorien. Nach Nikoloulopoulos und Karlis (2010) und anderen verwenden wir eine Copula, um die bivariate Wahrscheinlichkeitsmassefunktion von Oit und Yit wie folgt zu konstruieren:

|  |  |
| --- | --- |
|  | () |

wobei C(∙) die Copula-Funktion ist und F1(oit) und F2(yit) die Verteilungsfunktionen von Oit und Yit sind. In diesem Zusammenhang verwenden wir eine Frank-Copula-Funktion (z.B. Frank 1979; Genest 1987) aufgrund ihrer Flexibilität, den gesamten Bereich der Korrelation zu erfassen. Die Frank-Copula-Funktion ist gegeben durch:

|  |  |
| --- | --- |
|  | () |

Hierbei sind u1 und u2 die Verteilungsfunktionen (ZINBD und BD in unserem Fall) und q ist der Korrelationsparameter der Frank-Copula.

## Berichtigung der Endogenität

Es besteht die Möglichkeit, dass das Unternehmen Kunden gezielt anspricht, indem es je nach früheren Kaufaktivitäten und inhärenten Eigenschaften der Kunden (die wir nicht beobachtet haben) unterschiedliche Anzahlen von E-Mails sendet. Mit anderen Worten, die E-Mail-Kontaktvariable (d. h. EM) könnte endogen sein. Um dieses potenzielle Endogenitätsproblem in EM zu lösen, verwenden wir einen Kontrollfunktionsansatz (Petrin und Train 2010), der ähnlich wie der Ansatz von Villas-Boas und Winer (1999) ist. Dabei wird eine lineare Regression der EM auf Instrumente als erste Stufe durchgeführt, die Kunden-Kaufhäufigkeit und Kunden-Festeffekte umfasst, um andere Kundenmerkmale zu berücksichtigen, die vom Unternehmen beobachtet werden können (um gezieltes E-Mailing durchzuführen), jedoch nicht vom Forscher. Aus der ersten Schritt Regression erhalten wir ein R2 von 48,5%. Darüber hinaus stellen wir fest, dass die Koeffizienten der Kaufhäufigkeit und -häufigkeit positiv und statistisch signifikant sind. Im zweiten Schritt integrieren wir die Residuen aus der ersten Stufe Regression als zusätzliche Kontrollvariablen in die HMM-Übergangs- (Gleichung 3) und CPM-Komponenten (Gleichung 11) unseres vorgeschlagenen Modells für Kunden-E-Mail-Öffnungen und -Käufe.

## Model estimation

Es gibt vier Parameter-Sets, die aus unserem Modell geschätzt werden müssen: (1) Parameter der Übergangsmatrix (Gleichung 3), (2) Parameter der CEOM (Gleichungen 6-7), (3) Parameter der CPM (Gleichungen 9-12) und (4) der Korrelationsparameter der Frank-Kopula-Funktion (Gleichung 14). Gemäß Netzer, Lattin und Srinivasan (2008) schreiben wir den Vektor der bivariaten Wahrscheinlichkeitsmassefunktion als Diagonalmatrix Hit. Gegeben der vorgeschlagenen HMM-Struktur kann die Likelihood-Funktion für eine Sequenz von Beobachtungen [ (Y i1 = y i1 , Oi1 = o i1), ..., (Y it = y it, O it = o it) ] ausgedrückt werden:

|  |  |
| --- | --- |
|  | () |

wobei 1 ein S x 1 Vektor aus Einsen ist.

# Verwandte Arbeiten

Masterareit Lars

<file:///C:/Users/felix.oechslein/Downloads/Masterarbeit%20Lars%20Moestue.pdf>

Warum Messung des Costumer Engagment Levels für Email Marketing wichtig ist

Consistent with this argument, a study by Return Path (2015), an industry expert on email optimization, suggests that email frequency optimization should depend on customers’ engagement level. However, the study does not provide an actionable tool for firms to adopt to launch effective email marketing programs.

<http://returnpath.com/wp-content/uploads/2015/06/RP-Frequency-Report-FINAL.pdf>

Email Marketing als Weg Meschen von einem Produkt zu überzteugen

<https://journals.sagepub.com/doi/epub/10.1177/10949968221095552>

Literaturübersicht zu Big Data/Ai/ML und CRM

<file:///C:/Users/felix.oechslein/Downloads/10-1108_JBIM-07-2021-0332-1.pdf>

„The meaning and measurement of customer retention” -> insbesondere Aufschlüsselung des Wertes von Kundenloyalität

<file:///C:/Users/felix.oechslein/Downloads/palgrave.jt.5740035.pdf>

Bestes Buch zu CRM, viele wichtige Themen -> Kumar MVP!!!

<https://weblibrary.miu.edu.my/upload/ebook/management%20_and_business/2018_Book_CustomerRelationshipManagement.pdf>

Muss ich noch lesen -> Einordnung von Kundenbindung und Vorgaben zur Messung von Kundenbindung

file:///C:/Users/felix.oechslein/Downloads/s40547-017-0080-0.pdf

# Evaluation

## Datenauswahl

Grundsätzlich betrachten wir Daten von Oktober 2021 bis Dezember 2022. Abbildung 7 verdeutlicht, dass eine Betrachtung der Daten vor Oktober 2021 nicht repräsentati für die Kunden desThalia Theaters wäre, da durch Beschränkungen während der Corona Pandemie eine kaum repräsentative Azahl 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 3.

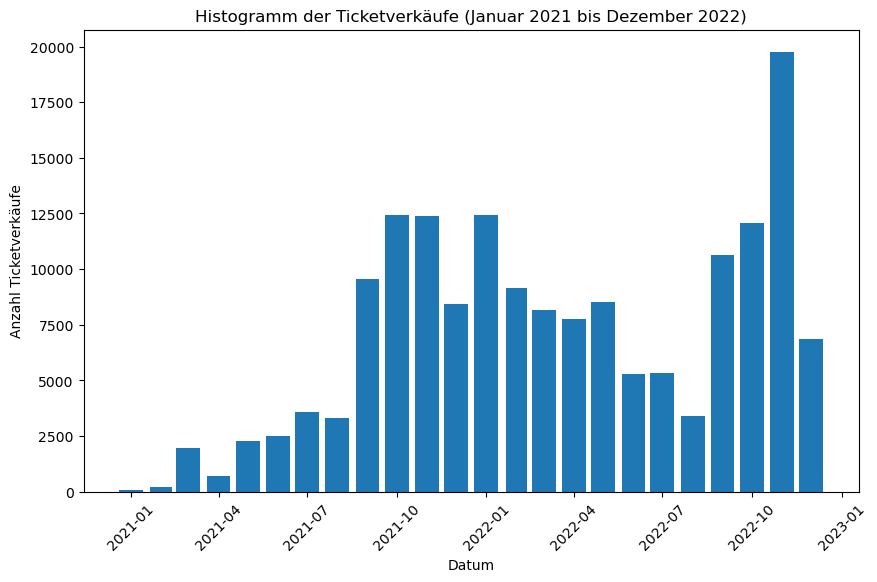


Abbildung : Histogramm derTicketverkäufe (Januar 2021 bis Dezember 2022)

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

**Privatkunden**

Es ist leicht verständlich, dass eine Messung der Auswirkugen der E-Mailkampagnen auf die Kundenbindung eine Einschräkung auf Privatkunden erfordert. Um diese Einschränkung durchführen zu können, bedarf es einer Metrik, die Privatkunden von Geschäftskunden unterscheiden 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, festgelgt. Nach Bereinigung der Trasaktionsdaten nach diesem Vorgehen, kann noch über 71% des Gesamtumsatzes abgebildet werden. Des Weiteren sind über der Hälfte der 30 größten Privatkunden und keiner der 30 größten Geschäftskunden nach händischer Überprüfung noch in dem Datenpool enthalten.

**Transaktionsdaten**

In dieser Arbeit werden Trasaktionen mit mehreren Tickets innerhalb einer Transaktion, als eine einzige Trasaktion 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 Trasaktion getätigt haben und nach 6 Moaten mindestens eine Weitere. Diese Einschräkung hat den Hintergrund, dass der Übergang von der Kudenloylität zur Kundenbindung das Untersuchungskriterium ist. Indem ausschließlich Kunden ausgewählt werden mit einer zusätzlichen Transaktion nach 6 Moaten, wird sichergestellt, dass z.B. Touristen und Kunden, die grundsätzlich unzufrieden mit der Leistung desTheaters waren, ausgeschlossen werden.

**Messung erhaltener Newsletter-E-Mails und E-Mailöffungen**

Bei der Auswahl der E-Mailkampagnen für die Berechnung findet eine Einschräkung 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 verwedet, die als "Newsletter" markiert sind. Wie in Kapitel 3.5.1 beschrieben, sind gemessene E-Mailöffungen der E-Mailprogramme 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.

## 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 (14) minimiert. Bei der Anwendung ist festzustellen, dass lediglich eine Anzahl on zwei Zuständen für die verwendeten Daten sinnvoll ist. Eine Auswertung mit drei oder mehr Zuständen hat zwar ähliche Wahrscheinlichkeiten erzielen können, aber hat immer einen der drei Zustände für sehr spezifische Kunden festgelegt und damit keine repräsentative 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 (<https://machinelearningmastery.com/bfgs-optimization-in-python/>).

**Startparameter**

Vor den Berechnungen wird zunächst der Dispersionparameter r der ZINBD-Verteilung geschätzt. Eine 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.

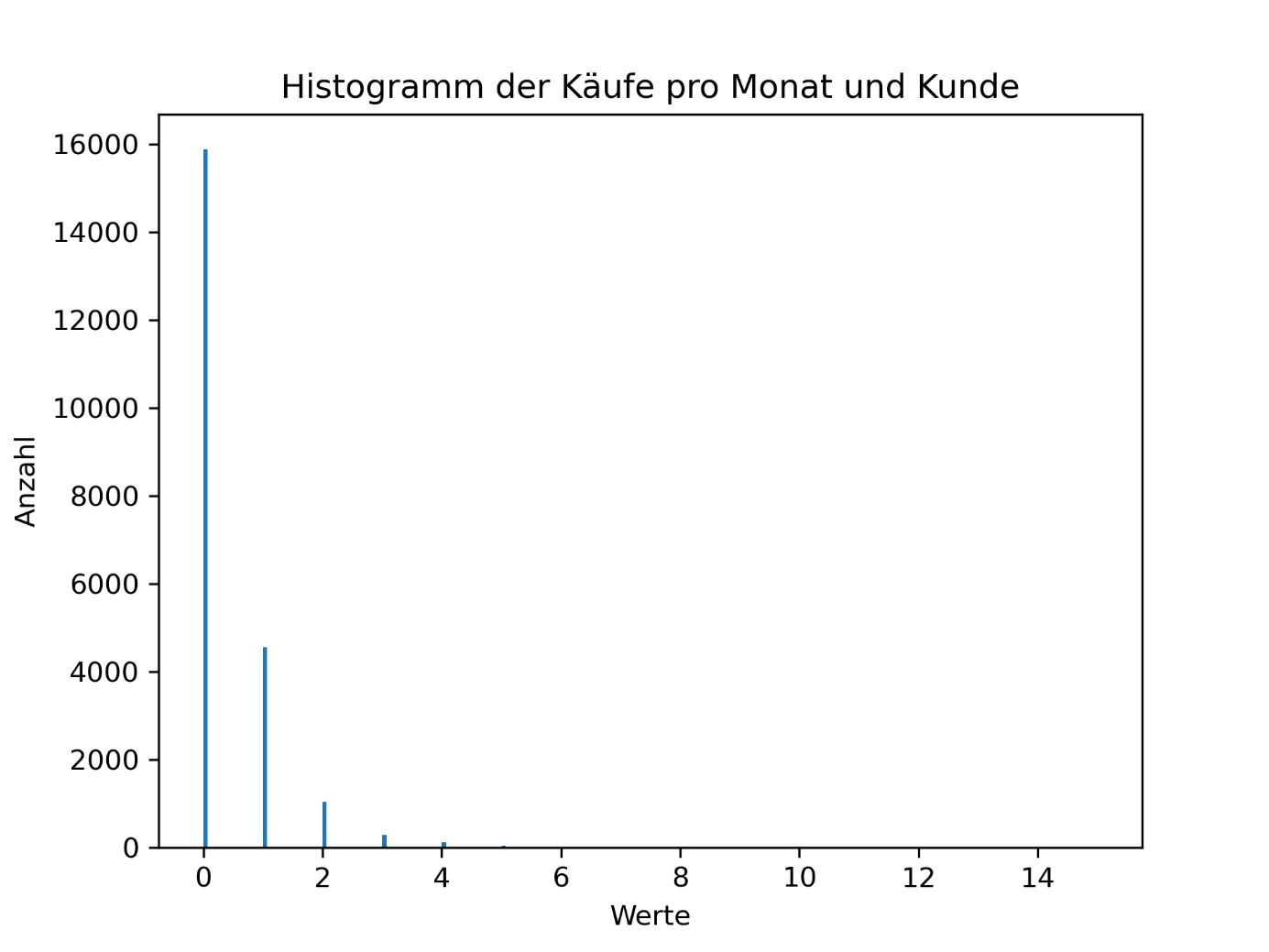


Abbildung : Histogramm der Käufe pro Monat und Kunde (im Zeitraum Oktober 2021 bis Dezember 2022)

Abbildung 8 zeigt das Histogramm der Azahl an Käufen pro Monat und Kunde. Aus diesen Daten wird die empirische Wahrscheinlichkeitsfunktion der Käufe ermittelt. Über die Minimierung der quadratischen Fehler der Wahrscheinlichkeitsfunktion einer ZINBD-, NBD und Poisson-verteilte Zufallsvariablen gegenüber den empirischen Wahrscheinlichkeitsfuktion für die gegebenen Werte werden die Parameter der jeweiligen Verteilung geschätzt. Für die ZINBD-Verteilung ergibt sich ein geschätzter Wert von 1.29 für den Dispersion-Parameter r.

Abbildung 9 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.

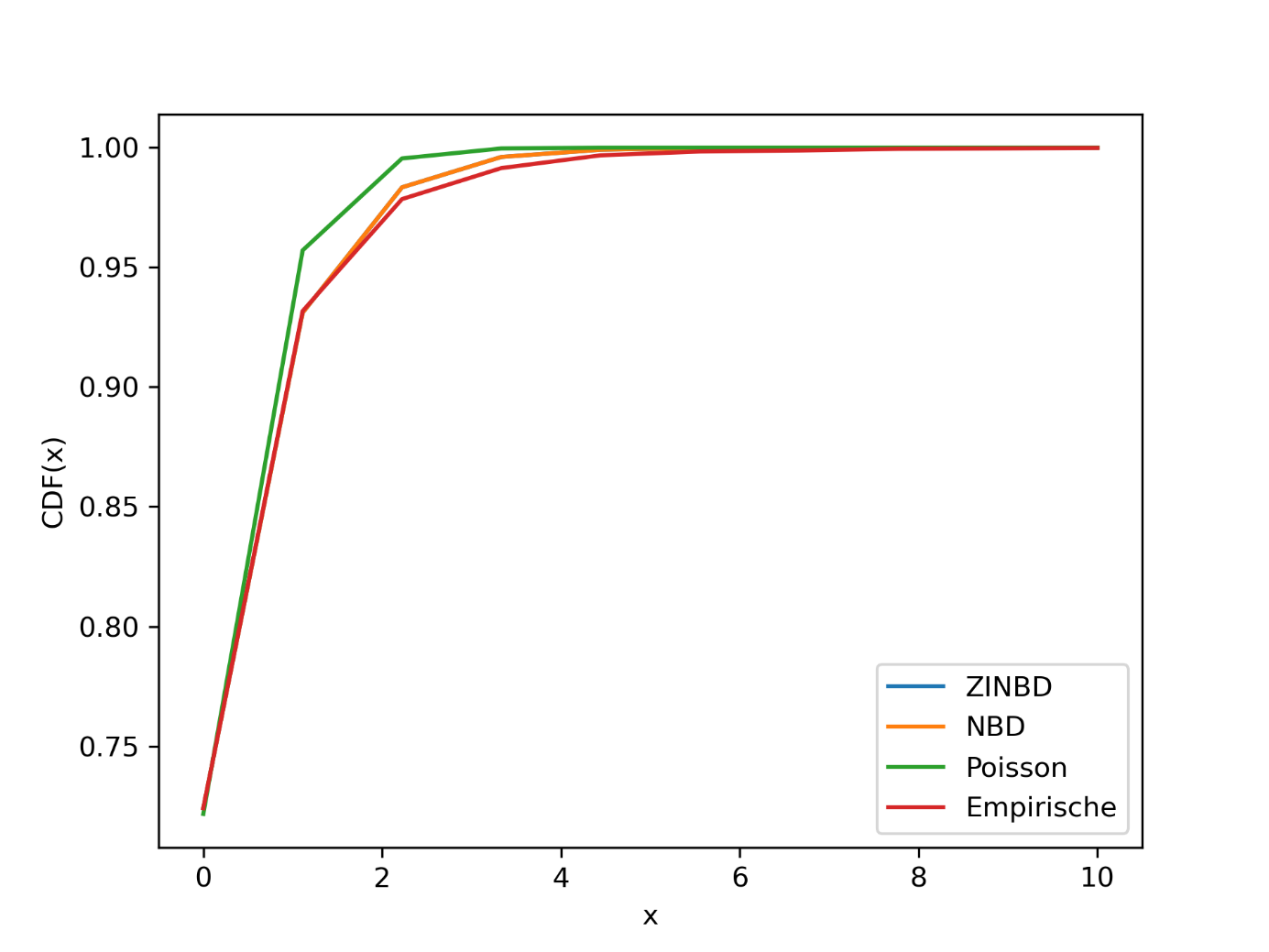


Abbildung : Vergleich der Verteilungsfunktionen (CDF) einer ZINBD-, NBD- und Poissonverteilten ZV mit geschätzten Parameterwerten gegenüber der empirischen Verteilungsfunktion

Da das Modell dazu genutzt wird, die Einflüsse des E-mailmarketings 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 : 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 untere Grenzen für die Parameter festgelegt.

## 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 Tabelle 2 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.

Formularbeginn

Formularende

|  |  |
| --- | --- |
| Parameter | Startwerte |
| alpha | [[0.0, -1.841],  [0.0, 0.990]] |
| gamma\_1 | [[0, 2.682],  [0, 2.806]] |
| gamma\_2 | [[0, -1.281],  [0, 0.698]] |
| gamma\_3 | [[0, 1.459],  [0, -2.130]] |
| gamma\_4 | [[0, -0.049],  [0, 0.622]] |
| alpha\_o | [0.331, 1.396] |
| beta\_o | [-1.121] |
| 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 (fest) | [1.29] |
| theta | [1.038] |

Tabelle : Parameterschätzungen (noch nicht aktualisiert)

**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 : Relevante Parameter für die Modellierung des Kaufverhaltens

Die Anzahl der Kundenkäufe wurde mittels einer ZINBD-Verteilung modelliert, wobei die Parameter delta\_0 und delta\_1 den Wert von φit|s in Gleichung (8) beeinflussen. φit|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 φit|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 delta\_0 und delta\_1 zu einem kleineren φit|s.

Nun wird der konkrete Einfluss der Parameter auf die Verteilung der Anzahl der Kundekäufe betrachtet. Derbeobachetete Wert LYit, 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 φit|2 Werte kleiner als 0.001 für LYit = 1, …, 14 annimmt.

In Zustand 1 wiederum kann für größere Werte von LYit, das heißt ein größerer Zeitraum seit dem letzen Kauf, ein zunehmender Einfluss der Zero-Inflation-Komponente betrachtet werden. Denoch kann auch im ersten Zustand der Einfluss der Zero-Inflation-Komponente vernachlässigt werden. So nimmt φit|1 Werte im Intervall [0.0001, 0.0109] für LYit = 1, …, 14 an.

Der Einfluss der Parameter alpha\_p, beta\_p\_1 und beta\_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 φit|s vernachlässigbar kleine Werte annimmt. Abbildung 7 zeigt den Einfluss von lambda auf die Wahrscheinlichkeitsfunktion der Käufe bei r = 1.29 und φit|s = 0 für i = 1, …, N, t = 1, …, T und s = 1, 2, 3.

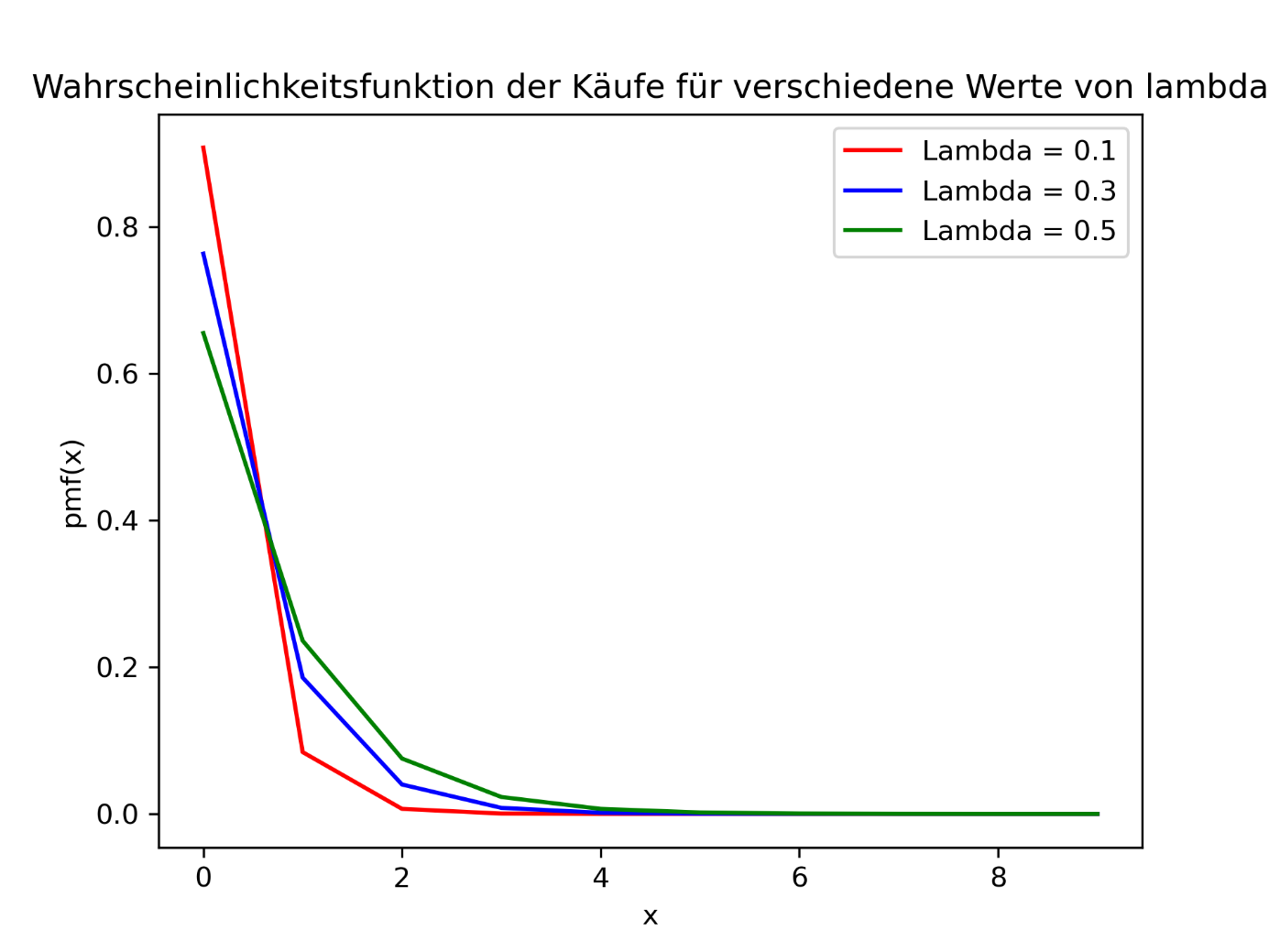
****

Abbildung : 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 an und für Kunden in Zustand 2 den Wert 0.182 an. Mit φit|s ≈ 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 : Einfluss der E-Mailkampagnen auf das Kaufverhalten

Tabelle 4 zeigt, dass Kunden in Zustand 1 weiger E-Mails bevorzugen. Da beta\_p\_2 für diese Kundengruppe negativ ist, ist der negative Einfluss quadratisch. Es gilt, je mehr E-Mails an Kundengruppe 1 gescickt werden, desto unwahrscheinliher ist deren Kauf. Kunden in Zustand 2 wiederum bevorzugen grundsätzlich mehr E-Mails zu erhalten, wobei der Einfluss des zusätzlichen E-Mailerhalts einen geringeren Einfluss auf deren Kaufverhalten hat.

**Interpretation der Parameter für die CEOM**

In Tabelle 4 sind die Schätzugen der relevanten Parameter für die Modellierung der E-mailöffungen 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) ≈ 2.9 mal mehr E-Mails als Kunden im Zustand 1. Beta\_O spiegelt den Einfluss der Zeit seit der letzten E-Mailöffung wider. Pro vergangenem Monat seit der letzten Öffnung sinkt die erwartete Anzahl der E-Mailöffnungen um ca. 32.5%.

|  |  |  |
| --- | --- | --- |
| Parameter | Zustand 1 | Zustand 2 |
| Alpha\_O | 0.331 | 1.396 |
| Beta\_O (zustandsinvariat) | -1.121 | |

Tabelle : Relevante Parameter für die Modellierung der E-Mailöffnungen

**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 ausgewählte Kunde 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 |
| *Keine erhaltene E-Mail* | 30.13% | 69.86% |
| *Eine erhaltene E-Mail* | 9.53% | 90.47% |
| *Zwei erhaltene E-Mails* | 2.76% | 97.24% |
| *Drei erhaltene E-Mails* | 0.83% | 99.16% |
| Von Zustand 2 |  |  |
| *Keine erhaltene E-Mail* | 2.20% | 97.80% |
| *Eine erhaltene E-Mail* | 9.22% | 90.78% |
| *Zwei erhaltene E-Mails* | 11.68% | 88.32% |
| *Drei erhaltene E-Mails* | 4.74% | 95.26% |

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

Tabelle 6 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 |
| *Keine erhaltene E-Mail* | 60.83% | 39.17% |
| *Eine erhaltene E-Mail* | 27.50% | 72.50% |
| *Zwei erhaltene E-Mails* | 9.27% | 90.73% |
| *Drei erhaltene E-Mails* | 2.94% | 97.06% |
| Von Zustand 2 |  |  |
| *Keine erhaltene E-Mail* | 1.10% | 98.90% |
| *Eine erhaltene E-Mail* | 4.81% | 95.19% |
| *Zwei erhaltene E-Mails* | 6.17% | 93.83% |
| *Drei erhaltene E-Mails* | 2.41% | 97.59% |

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

Tabelle 7 zeigt die Übergangsmatrix anhand eines Kunden mit einer Öffnung im letzten Monat und einem Kauf. Im Vergleich zu der Übergagsmatrix eines „typischen“ Kunden ohne Kauf im orherigen 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. Denoch hat die Azahl der versendeten Mails an diese Kundengruppe einen erheblichen Einfluss auf deren Verhalten im Modell. So kann die Wahrscheinlichkeit, dass sie den Zustand verlasse auf 97.06% erhöht werden bei drei erhaltenen Mails.

Des Weiteren ist der Einfluss des E-Mailerhalts auf die Übergangswahrscheinlichkeiten von Kunden im Zustand „mittlere Käufe“ deutlich geringer, falls sie im letzten Monat einen Kauf getätit haben. Dies lässt sich daran erkennen, dass dere Wahrscheinlichkeit im Zustand 2 zu bleiben für alle Anzahlen an versendete Mails größer ist alsei den „typischen“ Kunden. Denoch gilt auch hier, dass zwei E-Mails die optimale Azahl an E-Mails ist, um Kunden von 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 |
| *Keine erhaltene E-Mail* | 86.31% | 13.69% |
| *Eine erhaltene E-Mail* | 60.62% | 39.28% |
| *Zwei erhaltene E-Mails* | 29.31% | 70.69% |
| *Drei erhaltene E-Mails* | 10.97% | 89.03% |
| Von Zustand 2 |  |  |
| *Keine erhaltene E-Mail* | 27.09% | 72.01% |
| *Eine erhaltene E-Mail* | 62.68% | 37.32% |
| *Zwei erhaltene E-Mails* | 68.64% | 31.36% |
| *Drei erhaltene E-Mails* | 45.13% | 54.87% |

Tabelle : 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. Kunde, die breits 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, aber einen Kauf getätigt haben, siehe . Man erkennt, dass für diesen Fall die Wahrscheinlichkeit für Kunden in Zustand 1 in diesem Zustand zu bleiben maximiert wird. So haben diese Kunden beim Nicht-Erhalte von E-Mails eine 95.78% Wahrscheinlichkeit in ihrem Zustand zu bleiben. Ebenfalls können Kunden in Zustand 2 mit einer über 50% Erfolgswahrscheinlichkeit dazu bewogen werden, in den Zustand „viele Käufe“ zu wechseln beim Versenden von zwei E-mails an diese Kundengruppe.

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

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

## 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 (14) 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 denen 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:

|  |  |  |  |
| --- | --- | --- | --- |
| |  |  | | --- | --- | |  | () | |  |

ist die Wahrscheinlichkeit der beobachteten Sequenz von Entscheidungen bis zum Zeitpunkt t gemäß Gleichung (14). gibt die s-te Spalte der Matrix an. In Abbildung 11 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 Üergangswahrscheinlichkeit ist und nicht die getroffene Vorauswahl.

Es fällt auf, dass der Kundenbindungszustand große Schwakungen im Zeitverlauf aufweist. Selst Kunden, die sehr stark an das Unterehmen gebunden sind (gemessen an einer Wahrscheinlichkeit in Zustand 1 zu sein von konstant über 50%), weisen teilweise starke Schwakungen 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.

|  |  |
| --- | --- |
|  |  |
|  |  |

Abbildung : Wahrscheinlichkeiten von zufälligen Kunden in Zustand 1 zu sein

## 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 Unterhmens 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 abhägig davon, ob sie im vorherigen Monat eine Öffung 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 erhalteten E-mails festgestellt, siehe Tabelle 4. 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-mailerhalt und der Anzahl der Käufe aufgezeigt, siehe Tabelle 4. 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 Käufe“ bei einem E-Mailversand 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.

## Robustheitstests

Um zu überprüfen, ob unser Modell und unsere Ergebnisse robust gegenüber alternativen Modellspezifikationen und auf andere Kunden anwendbar sind, 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 Öffungszahlen
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-mailversendungen an Kunden in Zustand 1 und deren erwarteten Käufen, aber schätzt 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 erzeilt von 20927.46 im Vergleich zu 17936.24 mit dem in dieser Arbeit vorgestellten Modell. Abbildung 9 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 Berechung 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.

**Alterative Verteilung zur Modellierung der Öffnungszahlen**

Für die Verteilung der Öffungszahlen wurdeuntersucht, obe eine Negatie Binomialverteilung besser geeignet ist als die verwendete Poissonverteilung. Es ist bekannt, dass die Summe negativ binomialverteilter Zufallsvariablen wieder eine negativen Binomialverteilung folgen. Die Poissonverteilung ist ebenfalls reproduktiv. Zur Auswahl der geeigneten Verteilung wurde daher die empirische Verteilung der Öffungszahlen 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 Apassungen 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 Poisson-Verteilung als geeignetste Verteilung für die Modellierung bestimmt.

**Tests mit Vergleichskunden**

Des Weiteren wurde zur Validierung der Ergebisse 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 5.1 beschrieben, stehen noch Daten von 552 Kunden für die Valdierung zur Verfügung.

Zum Test der Modellspezifikationen wird getestet, ob eine Schätzung der Modellparameter für die Testkunden ebenfalls gute Einblicke in die Kundenbindun geben kann, oder ob beispielswiese eine Festlegung der Poissonverteilung für die Modellierung der Öffungen eine gute Anpassun des Modells nicht zulassen würde.

Zunächst werden die einzelnen Parameterschätzungen für die Testkunden mit den Parameterschätzungen der vorher verwendeten Kunden diskutiert. Die Parameterschätzungen zum Öffungsverhalten der Testunden sind sehr ähnlich zu den Schätzungen mit den in dieser Arbeit verwendeten Kunde. 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 einen 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 ähnliche absoluten logarithmierten Likelihood-Wert wie für die Ausgangsdaten. Mit den verwendeten Parametern wurde ein Likelihoodwert von 15274.15 erzielt. Das Modell kann durch Anpassung der Parameter bezüglich der Testkunden einen Likelihoodwert von 14632.94 erzielen. Dies deutet darauf hin, dass für eine bessere Schätzung der Parameter mehr Kunden und längere Betrachtungszeiträume verwendet werden müssen.

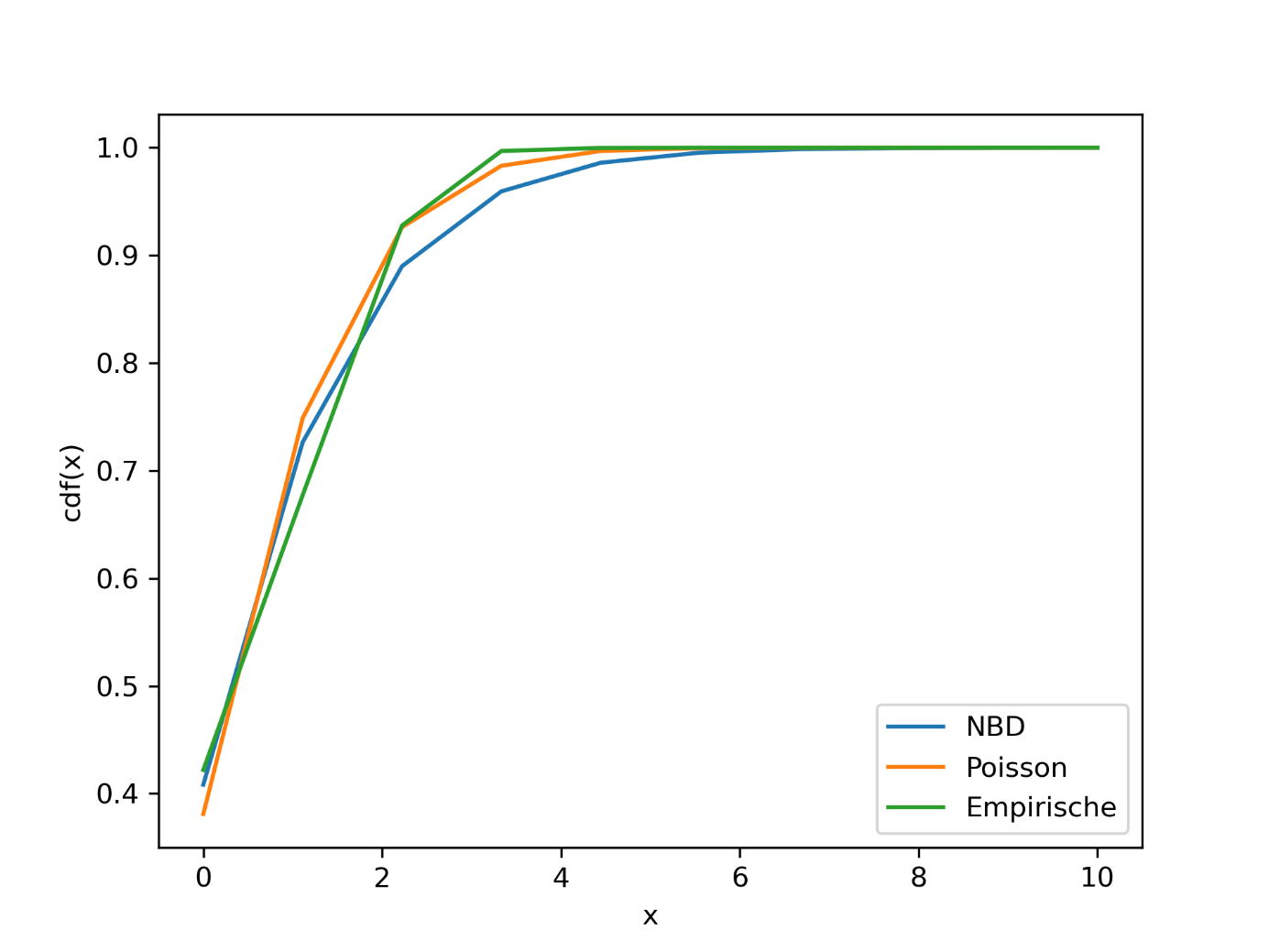


Abbildung : Vergleich der Poissonverteilung, NBD und der empirischen Verteilun der Öffungszahlen

# Optimales E-Mailmarketing

Die optimale Marketing-Kontaktstrategie steht im Mittelpunkt mehrerer früherer Studien (z. B. Khan, Lewis und Singh 2009; Kumar et al. 2011; Li, Sun und Montgomery 2011). In dieser Studie haben wir festgestellt, dass das Versenden von mehr E-Mails für das Unternehmen nicht unbedingt dazu führt, dass Kunden mehr Käufe tätigen. Daher kann die Bestimmung der optimalen Anzahl von E-Mails substantielle Gewinnzuwächse für das Unternehmen bieten. Unser Ziel in diesem Abschnitt besteht daher darin, ein DSS (Entscheidungsunterstützungssystem) bereitzustellen, auf das Manager vertrauen können, um den langfristigen Gewinn des Unternehmens zu maximieren, indem sie im Laufe der Zeit die optimale Anzahl von E-Mails an ihre Kunden senden.

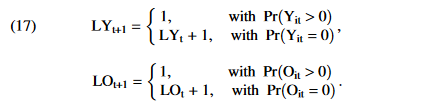
In dieser Situation muss das Unternehmen zu einem bestimmten Zeitpunkt t entscheiden, wie viele E-Mails gesendet werden sollen. Angesichts unserer geschätzten Kundenantwortfunktion hat die Entscheidung über den E-Mail-Kontakt sowohl kurz- als auch langfristige Auswirkungen auf das Kundenverhalten. Der kurzfristige Effekt ergibt sich aus der direkten Wirkung von E-Mails auf das Kundenkauf- und E-Mail-Öffnungsverhalten zum Zeitpunkt t (siehe Gleichungen 5 und 11).

Der langfristige Effekt ergibt sich aus zwei Quellen, nämlich (1) das Kauf- und E-Mail-Öffnungsverhalten der Kunden zum Zeitpunkt t beeinflusst die Entwicklung ihres Kundenbeziehungsstatus von Zeitpunkt t bis t + 1 (siehe Gleichung 3) und (2) die E-Mail-Kontakte zum Zeitpunkt t beeinflussen den Übergang des Beziehungsstatus von Zeitpunkt t bis t + 1 (siehe Gleichung 3). Aufgrund dieser langfristigen Effekte erfordert die Bestimmung der optimalen Anzahl von E-Mails, die vom Unternehmen gesendet werden sollen, die Lösung eines dynamischen Programmierungsproblems. Aus Sicht des Einzelhändlers ist die interessierende Variable die Anzahl der Käufe eines Kunden im Geschäft pro Monat. Unter der Annahme eines konstanten Kaufbetrags pro Transaktion und einer festen Bruttomarge für den Einzelhändler wird die Anzahl der Käufe direkt in den Gewinn des Unternehmens (pro Kunde und pro Transaktion) umgerechnet. Für das dynamische Optimierungsproblem des Unternehmens sind die relevanten Zustandsvariablen (1) die Wahrscheinlichkeiten, dass der Kunde in jedem der Kunden-Unternehmens-Beziehungszustände existiert (p1t, p2t, p3t = 1 - p1t - p2t), (2) die Zeit seit dem letzten Kauf (LY) und (3) die Zeit seit dem letzten Öffnen einer E-Mail (LO). Daher wird der Zustandsvektor zum Zeitpunkt t zu St = (p1t, p2t, LYt, LOt). Gemäß Kumar et al. (2011) nehmen wir die zeitliche Abfolge der E-Mail-Kontaktentscheidungen wie folgt an. Zu Beginn jedes Monats t prognostiziert das Unternehmen die Wahrscheinlichkeit, dass der Kunde in jedem der drei Beziehungszustände p1t, p2t und p3t existiert (p3t = 1 - p1t - p2t). Basierend auf den prognostizierten Werten p1t, p2t und p3t sowie LYt und LOt entscheidet das Unternehmen, wie viele E-Mails an den Kunden gesendet werden sollen. Wir verwenden eine multinomiale Logit-Share-Funktion, um die Zustandsmitgliedschaftswahrscheinlichkeiten (p1t, p2t und p3t) zu erfassen und sie mit zwei Parametern w1t und w2t wie folgt in Beziehung zu setzen:

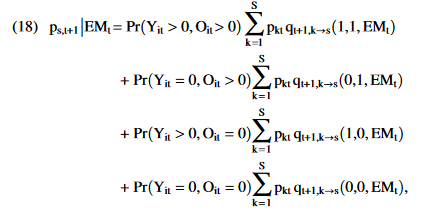
Ein Bild, das Text, Schrift, Screenshot, Reihe enthält.

Automatisch generierte Beschreibung

Sei St = (p1t, p2t, LYt, LOt) der Zustandsvektor zum Zeitpunkt t. Die Zeit seit dem letzten Kauf (LYt) und Öffnen (LOt) entwickeln sich auf der Grundlage davon, ob der Kunde zum Zeitpunkt t Käufe tätigt und E-Mails öffnet. Wenn der Kunde zum Zeitpunkt t Käufe tätigt, wird der entsprechende Zustand LYt+1 zu 1, andernfalls wird LYt+1 zu LYt + 1. Gleichzeitig wird, wenn der Kunde zum Zeitpunkt t E-Mails öffnet, der entsprechende Zustand LOt+1 zu 1, andernfalls wird LOt+1 zu LOt + 1. Da wir die Kauf- und E-Mail-Öffnungsprozesse der Kunden mit ZINBD und BD modellieren, entwickeln sich die Zustände der Zeit seit dem letzten Kauf und der Zeit seit dem letzten Öffnen, LYt und LOt, stochastisch wie folgt:



Die Entwicklung der ersten beiden Zustandsvariablen p1,t+1 und p2,t+1, abhängig von der E-Mail-Kontaktentscheidung des Unternehmens EMt, erfolgt gemäß folgender Gleichungen:



wobei qt+1,k→s(I(Y t > 0), I(O t > 0)), EMt ist die Übergangsfunktion aus Gleichung 4, die verwendet wird, um die Wahrscheinlichkeit zu berechnen, dass Kunden vom Zustand k zum Zeitpunkt t zum Zustand s zum Zeitpunkt t + 1 übergehen, unter Berücksichtigung von E-Mail-Kontakten, Kaufindikatoren und E-Mail-Öffnungsindikatoren zum Zeitpunkt t.

Zu jedem Zeitpunkt t, abhängig vom Zustandsvektor St, besteht das Ziel des Unternehmens darin, die optimale Anzahl von E-Mail-Kontakten zu bestimmen, um die summierten diskontierten zukünftigen Gewinne zu maximieren. Unter einigen Regularitätsbedingungen kann dieses Ziel in der folgenden Bellman-Gleichung formuliert werden:



wobei ρ der Diskontierungsfaktor ist, π(.) der Gewinn pro Periode ist und die Erwartung über alle zukünftigen Zustände und Aktionen des Unternehmens berechnet wird.

Um dieses dynamische Optimierungsproblem zu lösen, diskretisieren wir den Zustandsraum mit zehn Stufen für jede Zustandsdimension p1t , p2t , LYt und LOt, was zu 10.000 Zustandskombinationen führt. Wir verwenden den Value-Iteration-Algorithmus (Rust 1987), um die optimalen Zuordnungen der E-Mail-Kontakte des Unternehmens zu unseren gewählten Zustandskombinationen zu finden. Aufgrund der Diskretisierung der ersten beiden Zustandsdimensionen werden die Wertfunktionen für die anderen Punkte im Zustandsraum durch Interpolation berechnet (Keane und Wolpin 1994). Nachdem wir den Vektor der optimalen Zuordnung der E-Mail-Kontakte berechnet haben (durch den Value-Iteration-Algorithmus), passen wir ein multinomiales Logit-Modell (MNL) an, um optimale E-Mail-Kontakte als flexible Funktion des gewählten Zustandsraums vorherzusagen. Anschließend verwenden wir diese angepasste MNL-Richtlinienfunktion, um die optimale Anzahl von E-Mails für beliebige Zustandskombinationen aus dem gewählten Zustandsraum vorherzusagen.

Basierend auf der optimalen Zuordnung von E-Mail-Kontakten zu Zuständen aus dem Value-Iteration-Algorithmus stellen wir fest, dass die optimale Anzahl der E-Mail-Kontakte zwischen 5 und 14 liegt. Wir beobachten eine große Heterogenität in den Bereichen der optimalen Anzahl von gesendeten E-Mails basierend auf den Kundenbeziehungszuständen. Zum Beispiel, wenn das Unternehmen glaubt, dass die Wahrscheinlichkeit, dass der Kunde sich im Zustand mit hohen Käufen befindet, größer als 30% ist, liegt die optimale Anzahl von E-Mails zwischen fünf und sieben (basierend auf den Werten der verbleibenden Zustandskombinationen). Wenn das Unternehmen glaubt, dass der Kunde sich mit einer Wahrscheinlichkeit von mehr als 30% im Zustand mit mittleren Käufen befindet, liegt die optimale Anzahl von E-Mails zwischen sechs und zehn. Wenn das Unternehmen mit einer starken Überzeugung (mit einer Wahrscheinlichkeit von mehr als 90%) glaubt, dass der Kunde sich im Zustand mit niedrigen Käufen befindet, liegt die optimale Anzahl von E-Mails zwischen 12 und 14. Wenn das Gegenteil der Fall ist (d.h. das Unternehmen glaubt stark daran, dass der Kunde sich nicht im Zustand mit niedrigen Käufen befindet), liegt die optimale Anzahl von E-Mails zwischen fünf und sieben.

Diese Heterogenität in der optimalen Anzahl der zu sendenden E-Mails legt nahe, dass das Unternehmen verschiedene Kunden ansprechen sollte (indem es unterschiedliche Anzahlen von E-Mails sendet), basierend auf dem beobachteten Verhalten der Kunden (LY und LO) und der Überzeugung des Unternehmens über deren Beziehungszustände. Um zu veranschaulichen, wie das Unternehmen unser DSS zur gezielten Ansprache seiner Kunden über E-Mail-Marketing nutzen könnte, wählen wir zufällig zwei Kunden aus unserer Datenbank aus. Wir berechnen dann ihre HMM-Zustandsmitgliedschaftswahrscheinlichkeiten im Laufe der Zeit mithilfe des Filteransatzes und verwenden unsere optimale Richtlinienfunktion, um die Anzahl der E-Mail-Kontakte festzulegen (basierend auf ihren beobachteten LY- und LO-Niveaus in den Daten). Wir stellen die optimale Anzahl der E-Mail-Kontakte im Laufe der Zeit für diese beiden Kunden in Abbildung 5 dar. Wie Abbildung 5 zeigt, variiert die optimale Anzahl der zu sendenden E-Mails nicht nur zwischen diesen beiden Kunden erheblich, sondern auch beim selben Kunden im Laufe der Zeit. Diese Übung legt nahe, dass das Unternehmen unser vorgeschlagenes DSS leicht als handlungsorientiertes Managementinstrument verwenden kann, um seine E-Mail-Kontaktentscheidungen dynamisch und profitabel zu steuern.

Anschließend verwenden wir unsere optimale Richtlinienfunktion, um die E-Mail-Kontaktentscheidungen des Unternehmens zusammen mit den Kundenkauf- und E-Mail-Öffnungsreaktionen über einen langen Zeitraum für drei repräsentative Kunden in den Kategorien geringer, mittlerer und hoher Kaufzustände zu simulieren. Diese Vorwärtssimulation hilft uns zu verstehen, wie sich der Kundenlebenswert (CLV) bei Kunden in unterschiedlichen Kaufzuständen unterscheidet. In Übereinstimmung mit unseren Annahmen (siehe Fußnote 9) stellen wir fest, dass die CLVs von Kunden in den Kategorien geringer, mittlerer und hoher Kaufzustände jeweils 1.333 USD, 1.411 USD und 1.465 USD betragen. Diese Übung zeigt auch, dass in einem stabilen Zustand die optimale Anzahl von E-Mails, die an jeden Kunden gesendet werden sollen, sieben pro Monat beträgt. Darüber hinaus betragen die rabattierten Summen der Kauf- und E-Mail-Öffnungszahlen über die gesamte Kundenlebensdauer jeweils 73 bzw. 118 pro Kunde.

Schließlich führen wir eine "Was-wäre-wenn"-Simulationsstudie durch, um zu messen, wie viel Gewinn das Unternehmen verlieren würde, wenn es von der wiederhergestellten optimalen E-Mail-Richtlinienfunktion abweichen würde (siehe Abbildung 6). Wir verwenden die Verteilung im stabilen Zustand als Ausgangszustände und testen alternative Szenarien, in denen das Unternehmen von der optimalen Anzahl von sieben E-Mails abweicht. Wir testen die Szenarien, in denen das Unternehmen vier, fünf, sechs, acht, neun und zehn E-Mails anstelle von sieben sendet. Abbildung 6 zeigt, dass das Senden einer nicht optimalen Anzahl von E-Mails dazu führen kann, dass das Unternehmen einen erheblichen Gewinnverlust verzeichnet. Wenn zum Beispiel vier (zehn) E-Mails gesendet werden, verliert das Unternehmen 32% (16%) seines lebenslangen Gewinns pro Kunde. Dieses Ergebnis legt nahe, dass das Senden der optimalen Anzahl von E-Mails für die Rentabilität des E-Mail-Marketingprogramms des Unternehmens entscheidend ist.

# 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.

Optimales E-Mailmarketing Formularbeginn

Formularende

# Anlage

#### Anlage 1: Bezeichnung

Nachdem der Kernteil der Arbeit abgeschlossen wurde, wird wieder dieselbe Formatierung wie vor diesem verwendet:

* Schriftart Arial
* Text: Schriftgröße 11 pt
* Überschrift: Schriftgröße 11 pt, evtl. fett
* Zeilenabstand 1,5
* Ausrichtung: Blocksatz
* Silbentrennung wird empfohlen
* Seitenzahlen: römische Ziffern, anschließend an die des Vorspanns der Arbeit

Im Anhang können ergänzende Tabellen, Grafiken, Übersichten oder Interviews eingefügt werden. Auf weitere Erläuterungen wird an dieser Stelle allerdings verzichtet.

# 

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. kulturplanner; Theater Basel. kulturplanner, 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

Titel der Arbeit

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