ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Schemm, Jan; Schwarz, Christian; Strickrodt, Marc

Working Paper

Proaktives Kundenbindungsmanagement im Werbeartikelhandel: Entwicklung eines Machine-Learning-Modells zur Prognose von Kundenabwanderungen

Düsseldorf Working Papers in Applied Management and Economics, No. 60

Provided in Cooperation with:
Department of Business Studies, Hochschule Düsseldorf - University of Applied Sciences

Suggested Citation: Schemm, Jan; Schwarz, Christian; Strickrodt, Marc (2024) : Proaktives Kundenbindungsmanagement im Werbeartikelhandel: Entwicklung eines Machine-Learning-Modells zur Prognose von Kundenabwanderungen, Düsseldorf Working Papers in Applied Management and Economics, No. 60, Hochschule Düsseldorf, Fachbereich Wirtschaftswissenschaften, Düsseldorf, https://doi.org/10.20385/opus4-4633

This Version is available at:
https://hdl.handle.net/10419/306534

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Ausgabe 60
Oktober 2024

Hochschule Düsseldorf
University of Applied Sciences

Fachbereich
Wirtschaftswissenschaften
Faculty of
Business Studies

Forschungsberichte

des Fachbereichs Wirtschaftswissenschaften der Hochschule Düsseldorf

Jan Schemm, Christian Schwarz, Marc Strickrodt

Proaktives Kundenbindungsmanagement im Werbeartikelhandel: Entwicklung eines Machine-Learning-Modells zur Prognose von Kundenabwanderungen

Düsseldorf Working Papers in Applied Management and Economics

IMPRESSUM

Proaktives Kundenbindungsmanagement im Werbeartikelhandel:
Entwicklung eines Machine-Learning-Modells zur Prognose von Kundenabwanderungen

Autoren

Jan Schemm, Christian Schwarz, Marc Strickrodt

DOI: https://doi.org/10.20385/opus4-4633
URN: https://nbn-resolving.org/urn:nbn:de:hbz:due62-opus-46335

Lizenzhinweis: Es gilt das Urheberrecht.

Forschungsberichte des Fachbereichs Wirtschaftswissenschaften der Hochschule Düsseldorf
(Düsseldorf Working Papers in Applied Management and Economics)

ISSN: 2365-3361

Ausgabe: 60

Sämtliche Ausgaben dieser Reihe sind als Volltexte abrufbar unter:
https://wiwi.hs-duesseldorf.de/forschung/publikationen/forschungsberichte

Düsseldorf, Oktober 2024

Hochschule Düsseldorf
Fachbereich Wirtschaftswissenschaften
Münsterstr. 156
40476 Düsseldorf

Proaktives Kundenbindungsmanagement im Werbeartikelhandel: Entwicklung eines Machine-Learning-Modells zur Prognose von Kundenabwanderungen

Jan Schemm, Christian Schwarz

Hochschule Düsseldorf
Fachbereich Wirtschaftswissenschaften
Münsterstraße 156
40476 Düsseldorf
jan.schemm@hs-duesseldorf.de
christian.schwarz@hs-duesseldorf.de

Marc Strickrodt

WER GmbH
Messingweg 1
48308 Senden
marc.strickrodt@wer-gmbh.de

Abstract: Die Arbeit entwickelt systematisch ein Machine-Learning-Modell zur Prognose von Kundenabwanderungen im Werbeartikelhandel. Im Fokus steht die WER GmbH, ein mittelständischer Werbeartikelhändler, der jährlich signifikante Umsatzverluste durch Kundenabwanderung in der Streckenabwicklung verzeichnet und diese durch effektive Bindungsmaßnahmen reduzieren möchte. Die Ausgangsbasis für ein proaktives Kundenbindungsmanagement bildet ein Modell zur Identifikation abwanderungsgefährdeter Kunden. Das in einem Vergleich von insgesamt 15 Verfahren ausgewählte heterogene Machine-Learning-Ensemble nutzt eine Vielzahl transaktions-, leistungs-, kunden- und interaktionsbezogener Merkmale und liefert signifikant bessere Abwanderungsprognosen als einfachere Vergleichsverfahren. Zusätzlich zur inhaltlichen Interpretation des Modells und der relevantesten Merkmale beschreibt die Arbeit die praktische Integration in den Geschäftsablauf des Unternehmens. Sie liefert damit eine empirische Fallstudie zur Entwicklung eines Abwanderungsprognosemodells in nicht-vertraglichen B2B-Kundenbeziehungen und demonstriert die Leistungsfähigkeit datengetriebener Verfahren des maschinellen Lernens in der praktischen Anwendung.

1 Einleitung

Das Management von Kundenbeziehungen rückt für Unternehmen in vielen Branchen zunehmend in den Mittelpunkt ihrer Marketing- und Vertriebsaktivitäten. Insbesondere die Kundenbindung, d.h. die Verlängerung von profitablen Kundenbeziehungen, verspricht langfristig positive Wirkungen auf den ökonomischen Unternehmenserfolg.

Die WER GmbH, ein mittelständischer Werbeartikelhändler mit Hauptsitz im Münsterland, identifizierte Probleme und damit verbundene Chancen im Bereich seiner Kundenbindung. Erste Untersuchungen deuteten darauf hin, dass dem Unternehmen jährlich ca. 20-25% des Umsatzes in einem betrachteten Kundensegment durch Kundenabwanderung verloren gehen. Durch die Planung, Steuerung und Kontrolle von Kundenbindungsmaßnahmen soll diese Kundenabwanderung systematisch reduziert werden. Ziel des Unternehmens ist es, im Sinne einer proaktiven Prävention abwanderungsgefährdete Kunden zu identifizieren und diese durch geeignete Kundenbindungsmaßnahmen an das Unternehmen zu binden.

Zu diesem Zweck entwickelt und evaluiert die vorliegende Arbeit ein datengestütztes Modell zur Prognose von Kundenabwanderungswahrscheinlichkeiten auf Basis des maschinellen Lernens. Das Modell soll die kundenindividuelle Wahrscheinlichkeit für die Beendigung der Beziehung zum Unternehmen präzise prognostizieren und dabei in der Prognoseleistung signifikant bessere Ergebnisse als etablierte Vergleichsbenchmarks erzielen.

Der Aufbau der Arbeit unterteilt sich in sechs Teile. Der folgende Abschnitt 2 fasst den Stand der Forschung und der betrieblichen Praxis zum proaktiven Kundenbindungsmanagement mit Abwanderungsprognosen knapp zusammen. Nach einem Branchenüberblick beschreibt der Abschnitt Bedeutung und Verfahren der proaktiven Kundenbindung und arbeitet den Beitrag der Arbeit heraus. Abschnitt 3 stellt das betrachtete Unternehmen vor und der folgende Abschnitt 4 definiert die methodische Vorgehensweise der Arbeit. Abschnitt 5 beschreibt die systematische Entwicklung und Evaluation des Prognosemodells im Detail und liefert darüber hinaus Informationen zur organisatorischen und IT-seitigen Einbettung des Modells in der betrieblichen Praxis des Unternehmens. Abschnitt 6.2 fasst die erreichten Ergebnisse zusammen und liefert einen Ausblick auf weitere mögliche Arbeiten.

2 Proaktives Kundenbindungsmanagement unter Nutzung von Abwanderungsprognosen im Werbeartikelhandel

2.1 Werbeartikelhandel

Im Jahr 2023 investierten Unternehmen in Deutschland gemäß einer Erhebung von Nielsen Media Research ca. EUR 33,81 Mrd. in Werbung (vgl. Nielsen 2024). Von diesen Bruttoinvestitionen entfielen im gleichen Jahr ca. EUR 3,36 Mrd. auf Werbeartikel (vgl. GWW 2024). Mit einem Anteil von ca. 9,9% am Gesamtvolumen der Werbeausgaben spielt der Werbeartikel eine signifikante Rolle im Marketing- und Kommunikationsmix von deutschen Unternehmen.

Gemäß aktuellen Studien des Gesamtverbands der Werbeartikel-Wirtschaft setzen Unternehmen Werbeartikel hauptsächlich aufgrund ihrer vergleichsweise langen Werbewirkung (wiederholende Nutzung) und der positiven Abstrahlung auf das Unternehmensimage ein (vgl. GWW 2023). Rund zwei Drittel der Werbeartikelumsätze entfallen dabei auf Streuwerbeartikel mit einem Wert bis 10 €. Höherwertige Werbeartikel für mehr als 35 € machen weniger als 10% des Umsatzes aus (vgl. GWW 2024).

Trotz der auf den ersten Blick einfachen und günstigen „Give-Aways“ zeichnen sich Beschaffung, Lagerung und Distribution von Werbeartikeln durch eine gewisse betriebswirtschaftliche Komplexität aus. So ist bspw. eine Übersättigung der Kunden durch Nutzung gleichartiger Werbeartikel zu vermeiden und ein großer Teil der Artikel ist heute durch Sonderdesigns oder eine Werbeanbringung so individualisiert worden, dass die Beschaffungsabläufe entsprechend flexibel gestaltet werden müssen (vgl. Nebert 2006, S. 9). Darüber hinaus verursacht auch die Verteilung der Artikel an den jeweiligen Empfänger unter Berücksichtigung von Ansprüchen der beteiligten Abteilungen, Außendienstmitarbeiter oder Filialen in der Praxis einen erheblichen Arbeits- und Kostenaufwand für viele Unternehmen.

Aus diesen Gründen beziehen Marketing- und Vertriebsabteilungen in Unternehmen ihre Werbeartikel häufig nicht direkt vom Hersteller, sondern über spezialisierte Werbeartikelhändler. Die klassische Funktion des Werbeartikelhändlers besteht in der Vermittlung zwischen Herstellern bzw. Importeuren und dem beschaffenden Unternehmen. Er berät die Kunden – häufig individuell vor Ort – über Produktideen und Einsatzmöglichkeiten und übernimmt bei Abschluss die Produktion bzw. Beschaffung des Werbeartikels (vgl. Nebert 2006, S. 39 f.). Bei Kunden mit einer hohen Komplexität der Werbeartikelbewirtschaftung – bspw. charakterisiert durch viele verschiedene Artikel, eine hohe Umschlagshäufigkeit oder eine hohe Zahl an Abrufenden – zeigt sich ein Trend zur Auslagerung des gesamten Prozesses. Der Werbeartikelhändler übernimmt in diesem Fall sämtliche Aufgaben von der Produktvorstellung über den Einkauf, die Lagerhaltung und Logistik bis hin zum Reklamationsmanagement (vgl. Nebert 2006, S. 61 ff.).

Die Anbieterseite des Werbeartikelhandels in Deutschland ist mittelständisch geprägt und umfasst eine Größenordnung von 3´000 Unternehmen (vgl. GWW 2024). Ursprünglich örtlich lokal geprägt zeigt sich auch in dieser Branche eine Tendenz zur Digitalisierung: das Internet bildet heute die primäre Informationsquelle für Werbeartikel und Kunden setzen zunehmend mehr auf telefonische oder videobasierte Beratung (vgl. GWW 2023). Vor diesem Hintergrund und dem aus Kundensicht vermeintlich einfachen Wechsel des Beschaffungsweges wird es für Werbeartikelhändler zunehmend wichtig, ihre bestehenden Kunden an sich zu binden.

2.2 Proaktives Kundenbindungsmanagement

2.2.1 Bedeutung der Kundenbindung

Das Management von Kundenbeziehungen rückt für Unternehmen zunehmend in den Mittelpunkt ihrer Marketing- und Vertriebsaktivitäten (vgl. Bruhn 2022, S. 3; Kotler u. a. 2019, S. 59). In Abgrenzung zu einer transaktionsorientierten Marketing-Steuerung mit Instrumenten aus den traditionellen 4P-Dimensionen (Produkt, Preis, Promotion, Platzierung) des Marketingmix legt das Kundenbeziehungsmanagement den Fokus der Marketing-Instrumente auf in Abbildung 1 dargestellte Aufgabenbereiche entlang der Phasen einer Kundenbeziehung (vgl. Heinemann 2022, S. 104 f.; Stauss 2011, S. 333 f.):

- Die *Kundengewinnung* soll neue Kunden mit Ertragspotenzial und Profitbeitrag akquirieren.
- Die *Kundenbindung* soll die Beziehungen zu profitablen Kunden verlängern.
- Die *Kundenwertsteigerung* soll unprofitable bzw. umsatzschwache Kunden zu profitablen bzw. umsatzstarken Kunden migrieren.
- Die *Kundenrückgewinnung* soll Kunden, die die Geschäftsbeziehung verlassen haben, wiedergewinnen.

Von diesen Aufgabenbereichen verspricht insbesondere die Kundenbindung langfristig positive Wirkungen auf den ökonomischen Unternehmenserfolg (vgl. Bruhn 2022, S. 3 ff.).

Die Wirkungseffekte der Kundenbindung lassen sich durch theoretisch plausible Wirkungsketten auf der Erlös- und Kostenseite eines Unternehmens argumentieren. Erlöswirksam wird die Kundenbindung u.a. insofern, als dass sie die Preiselastizität der Nachfrage mindern und somit zu höheren realisierbaren Preisen führen kann: zufriedene Kunden bemühen sich weniger um einen Preisvergleich und schließen damit preisgünstigere Wettbewerber vom Geschäft aus (vgl. Bruhn 2022, S. 3; Diller 1995, S. 49; Reichheld 1999, S. 49 f.). Darüber hinaus ist die Wiederkaufwahrscheinlichkeit bzw. die Kauffrequenz gebundener Kunden per Definition höher, so dass der zu erwartende Lebenszyklusumsatz höher sein wird als im Fall ungebundener Kunden (vgl. Diller 1995, S. 39 ff.).

Zusätzlich kann die bestehende Kundenbeziehung zur Realisierung von Up- oder Cross-Selling-Potenzialen genutzt werden (vgl. Bruhn 2022, S. 3; Diller 1995, S. 39 ff.). Indirekt können dem Unternehmen verbundene Kunden den Erlös weiter steigern, indem sie das Unternehmen unter Umständen weiterempfehlen (vgl. Bruhn 2022, S. 4; Diller 1995, S. 42 f.; Reichheld 1999, S. 48 f.).

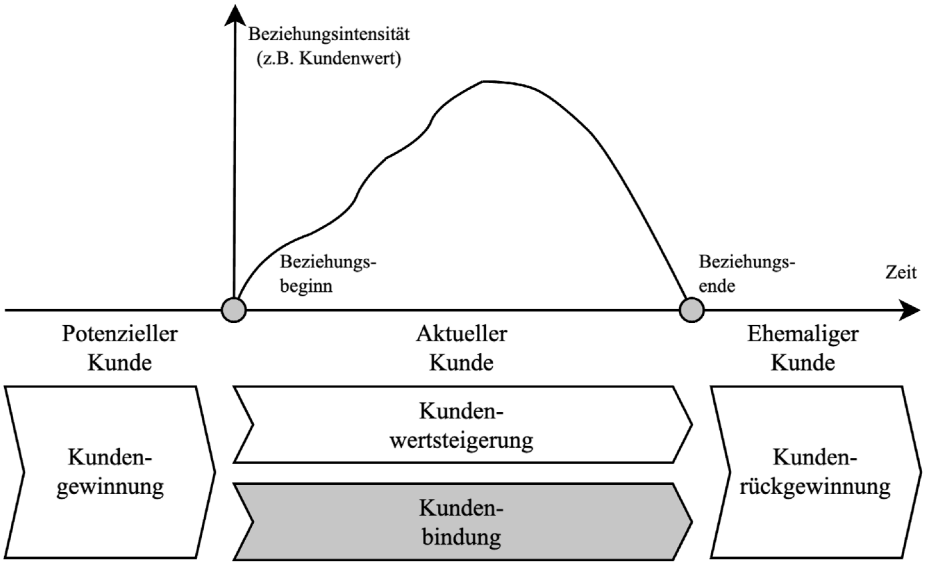

Abbildung 1: Aufgabenbereiche entlang des Kundenlebenszyklus (in Anlehnung an Boenigk 2011, S. 477; Stauss 2000, S. 16)

Auf der Kostenseite versprechen Erfahrungseffekte Kostensenkungspotenziale im Bereich der Bearbeitung gebundener Kunden im Vergleich zu neu gewonnen Kunden, bspw. durch geringere laufende Kosten der Kundenbearbeitung, effizientere Auftragsverfahren oder die Einsparung von Werbe- bzw. Akquisitionskosten (vgl. Bruhn 2022, S. 4; Diller 1995, S. 48; Reichheld 1999, S. 45 ff.).

Diese theoretisch hergeleiteten Argumente für positive Effekte der Kundenbindung werden durch Ergebnisse aus einzelnen Fallstudien oder Branchenuntersuchungen untermauert. So stellten bspw. Reichheld und Sasser bereits in den 90er Jahren in einer Studie unter US-amerikanischen Unternehmen aus verschiedenen Dienstleistungssektoren einen überproportional starken Zusammenhang zwischen dem Grad der Kundenbindung und der Höhe des Gewinns fest: eine Verminderung der Kundenabwanderung um 5% führte bei den betrachteten Unternehmen zu einer mittleren Gewinnsteigerung zwischen 25% und 85% (vgl. Reichheld, Sasser Jr. 1990, S. 109). Eine ähnliche Untersuchung von fünf US-amerikanischen Unternehmen kam zu dem Ergebnis, dass eine einprozentige Erhöhung der Kundenbindung zu einer Erhöhung des Kundenwerts zwischen 2,45% und 6,75% führt (vgl. Gupta u. a. 2004, S. 16). Der Einfluss der Kundenbindung auf den Kundenwert war in dieser Untersuchung deutlich höher als bspw. der Einfluss der Kundenakquisitionskosten oder der Margen.

2.2.2 Prognosefähigkeit als Voraussetzung für proaktive Kundenbindung

Kundenbindung und -rückgewinnung unterscheiden sich im Kern durch die zeitliche Terminierung der jeweiligen unternehmerischen Maßnahmen. Im Fall der Kundenrückgewinnung "wartet" ein Unternehmen, bis ein Kunde abgewandert ist und versucht diesen anschließend - i.d.R. mit finanziellen Anreizen - zurückzugewinnen. Im Fall der Kundenbindung agiert das Unternehmen im Sinn einer *Kündigungsprävention* (vgl. Boenigk 2011, S. 478) hingegen proaktiver und ergreift Maßnahmen um die Abwanderung von Kunden im Vorhinein zu verhindern. Tabelle 1 stellt beiden Ansätze anhand ausgewählter Vergleichskriterien gegenüber (vgl. Ascarza u. a. 2018, S. 75 ff.).

Tabelle 1: Kundenbindung vs. -rückgewinnung (in Anlehnung an Ascarza u. a. 2018, S. 75)

	Proaktive Kundenbindung	**Reaktive Kundenrückgewinnung**
Analytische Fähigkeiten	Prognosefähigkeiten notwendig	Beschreibende Analysen ausreichend
Genauigkeit der Ansprache	Risiken von Falsch-Positiven („schlafende Hunde") und Falsch-Negativen (Verpassen von Abwanderern)	Nur Kunden, die bereits abgewandert sind (tatsächlich realisierte Abwanderungen)
Wirksamkeit	Höhere Erfolgswahrscheinlichkeit, früh im Entscheidungsprozess	Kunde hat sich bereits entschieden, spät im Entscheidungsprozess
Kosten	Gewöhnlich geringere Anreize aufgrund Ungenauigkeit / Risiken der Ansprache	Gewöhnlich stärkere Anreize aufgrund präziser Möglichkeit der Ansprache und bereits getroffener Entscheidung
Nachhaltigkeit	Wahrscheinlicher unter der Annahme einer dauerhaften Wirkung auf Loyalität	Unwahrscheinlicher, Risiko strategischer Kündigungen

Maßnahmen zur Kundenrückgewinnung setzen keine Prognosefähigkeiten voraus und erzielen automatisch eine hohe Zielgenauigkeit: sie adressieren nur Kunden, die bereits abgewandert sind. Allerdings werden im Fall von Rückgewinnungskampagnen Kunden angesprochen, die ihre Entscheidung bereits getroffen haben. Diese Kunden wieder zum Verbleiben in der Geschäftsbeziehung bewegen zu können ist relativ „schwierig" bzw. erfordert einen vergleichsweise teuren Anreiz. Langfristig erscheinen reaktive Maßnahmen darüber hinaus wenig nachhaltig, da Kunden die Kampagnen mit der Zeit erwarten und damit das Risiko strategischer Kündigungen steigt.

Proaktive Kundenbindungsmaßnahmen erfordern im Vergleich zur Kundenrückgewinnung eine Investition in Prognosefähigkeiten: als erste Aufgabe der Kampagnenabwicklung sind zunächst abwanderungsgefährdete Kunden über Frühwarnindikatoren

oder statistische Modelle zu identifizieren. Selbst im Fall ausgeprägter Prognosefähigkeiten leidet die Zielgenauigkeit der Kampagnen aufgrund der Risiken von Fehlprognosen. So werden i.d.R. eine Reihe von abwanderungsgefährdeten Kunden nicht identifiziert (falsch negative Prognosen) und es kann zum Wecken von „schlafenden Hunden" kommen: d.h. Kunden, die keine Abwanderungsabsichten hegten, werden mit Anreizen zum Bleiben incentiviert. Auf der anderen Seite werden Kunden früher in ihrem Entscheidungsprozess bearbeitet, so dass mit einer höheren Wirksamkeit und Nachhaltigkeit der Kampagnen bei geringerem Einsatz an Anreizen gerechnet werden kann.

Proaktive Kundenbindung kann durch Planung, Steuerung und Kontrolle von passenden Maßnahmen umgesetzt werden. Abbildung 2 illustriert fünf Aufgabenbereiche im Rahmen der Abwicklung einer Kundenbindungsmaßnahme (vgl. Ascarza u. a. 2018, S. 70 ff.; Pick 2016, S. 61 ff.).

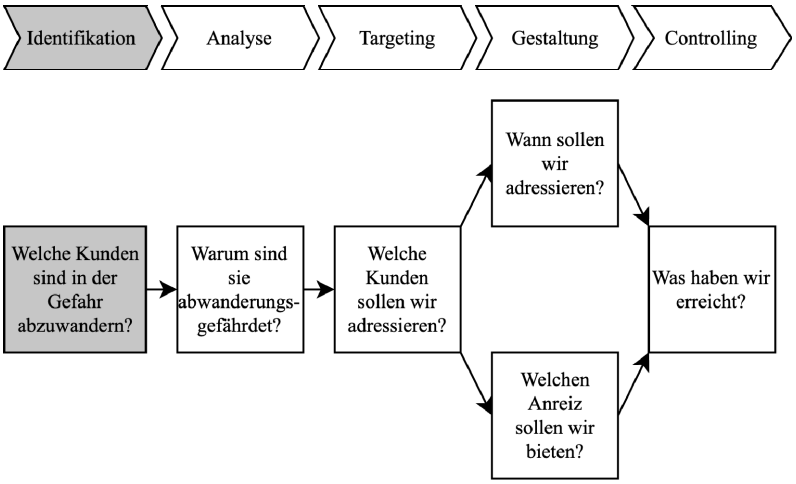

Abbildung 2: Aufgabenbereiche einer Kundenbindungsmaßnahme (in Anlehnung an Ascarza u. a. 2018, S. 70)

In einem ersten Schritt sind diejenigen Kunden zu identifizieren, die in der Gefahr sind abzuwandern (Identifikation). Im Zentrum steht hierbei die Ermittlung der potenziellen Gefährdung der Kundenbeziehung mit Hilfe von Frühwarnindikatoren oder statistischen Methoden. Danach sollten die Ursachen für das Abwanderungsrisiko analysiert werden (Diagnose). Die Aufgabe besteht darin, den zentralen Abwanderungsgrund bzw. das Zusammenspiel bestimmter Abwanderungsgründe zu ermitteln. Basierend auf dieser Diagnose ist festlegen, welche ausgewählten Kunden zur Vermeidung der potenziellen Abwanderung mit spezifischen Maßnahmen bearbeitet werden sollen (Targeting). Für die Auswahl der zu adressierenden Kunden sind insbesondere Kundenwertanalysen zielführend, die den langfristigen Ergebnisbeitrag von Investitionen in Bindungsmaßnahmen spezifischer Kunden beurteilen. Der Schritt danach bestimmt die konkreten Maßnahmen, d.h. legt fest mit welchen Elementen der Preis-, Produkt, Kommunikations- und Distributionspolitik die ausgewählten Kunden wann adressiert werden sollen (Ge-

staltung). Nach der Maßnahmendurchführung schließlich sollten der Erfolg bewertet und der Prozess weiterentwickelt werden (Controlling).

Die Gestaltung von proaktiven Kundenbindungsmaßnahmen setzt eine Identifikation von abwanderungsgefährdeten Kunden *vor* ihrer Abwanderung voraus, d.h. eine elementare Aufgabe besteht in der Entwicklung einer Prognose zur Kundenabwanderung.

2.3 Verfahren zur Kundenabwanderungsprognose

Prognosen zur Kundenabwanderung werden durch verschiedene Verfahren ermöglicht. Ein erstes, fundamentales Kriterium zur Klassifikation solcher Prognoseverfahren betrifft die Art der Beziehung zwischen Unternehmen und Kunden. Zu unterscheiden sind hierbei vertragliche und nicht-vertragliche Kundenbeziehungen (vgl. Fader, Hardie 2009, S. 62 ff.; Schmittlein, Peterson 1994, S. 16 f.). Im Fall einer vertraglichen Kundenbeziehung ist die Beendigung derselben i.d.R. eindeutig erkennbar, bspw. im Fall einer Vertragskündigung oder des Auslaufens eines Vertrags. Im Fall einer nicht-vertraglichen Kundenbeziehung hingegen kann der Zeitpunkt, zu dem ein Kunde inaktiv wird, vom Unternehmen nicht beobachtet werden. Kunden teilen dem Unternehmen in diesem Fall nicht mit, dass sie „abwandern", stattdessen stellen sie die Beziehung „stillschweigend" ein. Prognoseverfahren müssen die beiden genannten Fälle unterscheiden, d.h. i.d.R. ist es unangemessen ein für einen vertraglichen Kontext entwickeltes Verfahren in einem nichtvertraglichen Kontext anzuwenden und umgekehrt (vgl. Fader, Hardie 2009, S. 63).

Eine zweite Unterscheidungsdimension besteht in der wissenschaftlichen Kultur, in der die Verfahren ihren Ursprung haben bzw. dem Ausgangspunkt oder Treiber der Analyse (vgl. Breiman 2001a, S. 199 ff.; Januschowski u. a. 2020, S. 168 ff.). Modell-getriebene Verfahren auf der einen Seite beginnen die Analyse mit der Annahme einer bestimmten Form bzw. Verteilung der zugrundeliegenden Daten und bestimmen im Anschluss einzelne Parameter zur Erzielung einer guten Prognose. Datengetriebene Verfahren hingegen treffen a priori keine strukturellen Annahmen, sondern extrahieren Muster direkt aus den vorhandenen Daten. Sie sind i.d.R. flexibler als modell-getriebene Verfahren, benötigen jedoch häufig eine umfangreichere Datenmenge, um gute Ergebnisse zu erzielen. Modell-getriebene Verfahren werden häufig auch etwas allgemeiner als statistische Verfahren bezeichnet, während datengetriebene Verfahren häufig unter dem Begriff Machine Learning (ML) subsumiert werden. Diese Arbeit verwendet im Folgenden beide Begriffspaare jeweils synonym.

In einer dritten Dimension schließlich lassen sich alleinstehende und kombinierte Verfahren, sog. Ensembles, unterscheiden. Ensembles kombinieren mehrere alleinstehende Prognoseverfahren um die Prognoseleistung insgesamt zu verbessern (vgl. Bogaert, Delaere 2023, S. 2 f.; Rokach 2019, S. 51). Sie lassen sich weiter unterteilen in homogene und heterogene Ensembles (vgl. Bogaert, Delaere 2023, S. 2 f.). Homogene Ensembles bündeln die Prognosen einer Menge von gleichen, häufig eher einfachen Basis-Verfahren wie bspw. Entscheidungsbäumen. Diversität zwischen den einzelnen Prognosen wird dabei i.d.R. durch eine Partitionierung der Daten oder des Merkmalsraums erzeugt. Zwei bekannte Ansätze hierfür sind *Bagging* oder *Boosting* (vgl. Rokach

2019, S. 54 ff.). Im Vergleich dazu kombinieren heterogene Ensembles mehrere unterschiedliche Prognoseverfahren. Sie erzeugen Diversität primär über die verschiedenen Ansätze der Verfahren und wählen Kombinationen über Optimierungsverfahren aus. Die einzelnen Basisverfahren können dabei sowohl aus alleinstehenden Verfahren als auch aus homogenen Ensembles bestehen. Verbreitete Ansätze für die Kombination heterogener Prognoseverfahren sind *Voting* oder *Stacking* (vgl. Rokach 2019, S. 105 ff.).

Die vorliegende Arbeit entwickelt und bewertet Prognoseverfahren in einem nichtvertraglichen Kontext. Der Fokus der Analysen liegt dabei auf datengetriebenen, mehrstufigen Ensemble-Verfahren, da diese im Anwendungsgebiet die besten Prognoseleistungen versprechen (vgl. Bogaert, Delaere 2023, S. 22 f.). Alleinstehende und statistische Verfahren werden als Vergleichsmaßstäbe (Baselines) verwendet, um den Mehrwert der komplexeren Verfahren im konkreten Anwendungsfall zu beurteilen.

2.4 Beitrag der Arbeit

Die vorliegende Arbeit liefert neuartige Ergebnisse in einer Reihe von Bereichen. Zum einen entwickelt die Arbeit konstruktiv ein Prognosemodell, welches einen klar erkennbaren Beitrag zur Lösung eines praktisch relevanten Problems für das betrachtete Unternehmen liefert.

Die Arbeit liefert darüber hinaus eine empirische Fallstudie zur Entwicklung eines Abwanderungsprognosemodells für ein mittelständisches Unternehmen im Business-to-Business (B2B)-Bereich. Sowohl der B2B-Bereich als auch der Mittelstand erscheinen in bisherigen Arbeiten zu Abwanderungsprognosen unterrepräsentiert und den Autoren ist keine vergleichbare Arbeit in der Werbeartikelbranche bekannt.

Die Leistungsbewertung der entwickelten Modelle beschränkt sich nicht wie häufig in der Literatur auf den Metrik-basierten Vergleich verschiedener Algorithmen miteinander, sondern vergleicht die Modellleistung mit praktisch relevanten, einfachen Vergleichsverfahren und quantifiziert damit den Mehrwert der datengetriebenen Modellbildung.

Die Modell-Implementierung basiert ausschließlich auf frei verfügbaren Bibliotheken in der Programmiersprache Python und kann damit vergleichsweise einfach auf analoge Problemstellungen übertragen werden.

Schließlich endet die Arbeit nicht mit einer theoretischen Bewertung des entwickelten Modells, sondern beschreibt auch die organisatorische und IT-seitige Implementierung der Abwanderungsprognose in der betrieblichen Praxis, wo sich deren ökonomische Wirkung erst entfalten kann.

3 Unternehmen

Die WER GmbH ist ein mittelständischer Werbeartikelhändler mit Hauptsitz in Senden im westfälischen Münsterland (WER 2024). Das Unternehmen bietet seit 1993 Full-Service-Dienstleistungen im Handel mit Werbeartikeln an und beschäftigt aktuell 105 Mitarbeiter am Unternehmenssitz sowie weiteren 7 Vertriebs- und Logistik-Standorten in Deutschland. Durch organisches Wachstum sowie die strategische Übernahme von Marktbegleitern konnte das Unternehmen seinen Umsatz auf insgesamt ca. 18 Mio. € in 2023 steigern. Seit Juli 2024 gehört das Unternehmen zu Geiger, einem der weltweit größten familiengeführten Werbeartikelunternehmen mit Hauptsitz in den USA (Geiger 2024).

In Ergänzung zur klassischen Beschaffung von Werbeartikeln über eine Streckenabwicklung positioniert sich die WER GmbH strategisch als Outsourcing-Anbieter im Full-Service-Geschäft und bietet seinen Kunden die komplette Bandbreite des Dienstleistungsspektrums rund um den Werbeartikel von der Sortimentszusammenstellung über die Beschaffung der Produkte, die Lagerhaltung und Logistik, die integrierte Auftragsabwicklung über kundenindividuelle Webshop-Systeme bis hin zum Controlling.

Die Full-Service-Kunden trugen 2023 ca. 45 % zum Gesamtumsatz bei, während die verbleibenden ca. 55 % über die Streckenabwicklung erwirtschaftet wurden (vgl. Abbildung 3).

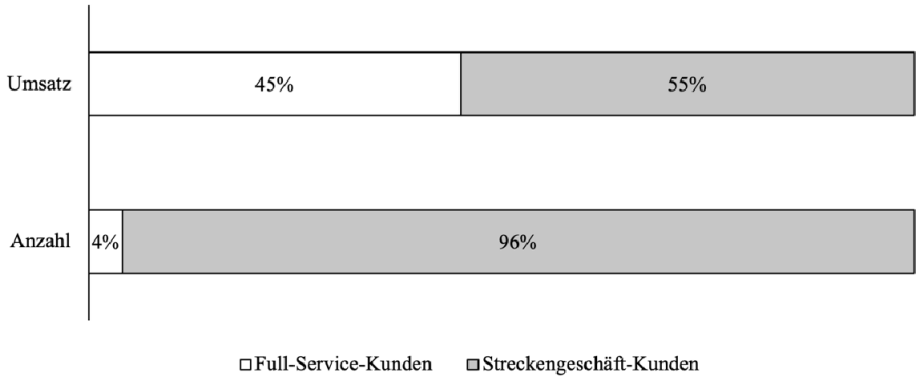

Abbildung 3: Kunden- und Umsatzverteilung 2023

Auslöser für die Beschäftigung mit einer proaktiveren Kundenbindung waren Beobachtungen im Streckengeschäft. Während die vertraglich organisierten Beziehungen zu Full-Service-Kunden sowohl aufgrund der vergleichsweise geringen Anzahl als auch der relativen Umsatzbedeutung eng geführt werden, ist die zielgerichtete Entwicklung von Kundenbeziehungen in der Streckenabwicklung deutlich anspruchsvoller. Erste Untersuchungen deuteten darauf hin, dass dem Unternehmen jährlich ca. 20-25% des Umsatzes durch die Abwanderung von Streckenkunden verloren gehen. Dieser Kunden- und Um-

satzverlust muss vor dem Hintergrund der Wachstumsziele des Unternehmens regelmäßig durch die vergleichsweise aufwändige Akquise von Neukunden kompensiert werden.

Durch die Planung, Steuerung und Kontrolle von proaktiven Kundenbindungsmaßnahmen soll die Kundenabwanderung im Streckengeschäft systematisch reduziert werden. Ausgangspunkt für diese Initiative ist die Entwicklung eines präzisen, datengestützten Prognosemodells.

4 Methode

Die Vorgehensweise zur Entwicklung des Modells zur Kundenabwanderungsprognose folgt der Methode *Cross Industry Standard Process for Data Mining* (CRISP-DM) (vgl. Chapman u. a. 2010; Helander 2021, S. 112 f.; Wirth, Hipp 2000), einem de facto Standard zur Entwicklung von Data-Mining-Lösungen (vgl. Mariscal u. a. 2010, S. 146). Die Entwicklung des Modells unterteilt sich in sechs Phasen (vgl. Abbildung 4):

1. Die Phase *Geschäftsverständnis* legt die Anforderungen und Ziele des Projekts aus einer betriebswirtschaftlichen Perspektive fest und überführt diese in eine analytische Problembeschreibung mit entsprechenden Zielen.

2. Die Phase *Datenverständnis* beginnt mit der Zusammenstellung für die Problemlösung relevanter Daten und analysiert die Daten im Hinblick auf die Zielsetzung.

3. Die Datenaufbereitung umfasst alle notwendigen Aufgaben zur Erstellung des Datensatzes als Eingabe für die Modellierung. Diese beinhalten die Auswahl und Integration von Daten für die folgenden Analysen sowie die Konstruktion von Merkmalen. Merkmale werden im Anschluss ausgewählt und im Hinblick auf die Anforderungen der verwendeten Modelle formatiert und ggf. transformiert.

4. In der *Modellierung* wird zunächst ein einheitliches Testdesign zur Bewertung verschiedener Modellierungstechniken festgelegt. Im Anschluss werden verschiedene Modellierungstechniken ausgewählt, kalibriert und anhand des Testdesigns bewertet.

5. Die *Evaluation* vergleicht die bewerteten Modelle und beurteilt, inwieweit das beste Modell die in der ersten Phase definierten Geschäftsziele erreicht. Darüber hinaus interpretiert die Phase das Modell, d.h. sie versucht die Funktionsweise des Modells zu erklären.

6. Die letzte Phase der *Bereitstellung* integriert das Modell in die Organisation und IT-Landschaft des betrachteten Unternehmens.

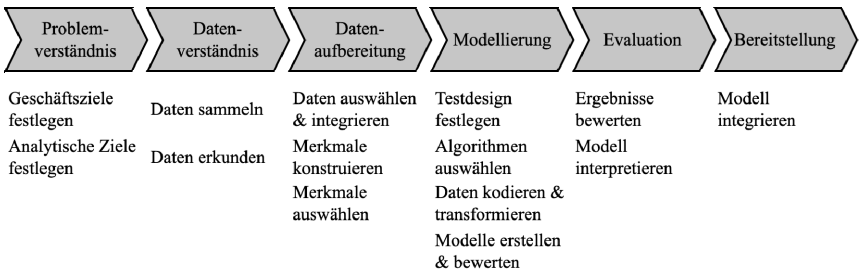

Abbildung 4: Vorgehen zur Entwicklung des Kundenabwanderungsprognosemodells

5 Modellentwicklung

5.1 Geschäftsverständnis

5.1.1 Geschäftsziele festlegen

Vor dem Hintergrund der in Abschnitt 3 beschriebenen Herausforderung verfolgt die WER GmbH das Ziel, bestehende Streckenkunden stärker an sich zu binden. Hierzu sollen abwanderungsgefährdete Kunden durch ein Modell frühzeitig identifiziert werden, um diese durch passende Maßnahmen zur Fortsetzung der Geschäftsbeziehung bewegen zu können.

5.1.2 Analytische Ziele festlegen

Problemformulierung

Übersetzt auf eine analytische Ebene möchte das Unternehmen monatlich Abwanderungswahrscheinlichkeiten für alle bestehenden Kunden prognostizieren um anhand dieser Bewertung – und ggf. weiterer Faktoren wie einer Kundenwertanalyse – Kunden für passende Bindungsmaßnahmen auswählen zu können.

Die Formulierung des Problems der Prognose von Abwanderungswahrscheinlichkeiten unterteilt die historischen Daten anhand eines gewählten Betrachtungszeitpunkts in zwei Zeitfenster (vgl. Abbildung 5):

- Das *Kalibrierungsfenster* umfasst Daten zum Kundenverhalten in einem definierten Zeitraum und dient der Ableitung informativer Eigenschaften zu jedem Kunden.

- Das *Prognosefenster* umfasst Daten zum Kundenverhalten in einem definierten anschließenden Zeitraum und dient der Beobachtung des interessierenden Kundenverhaltens, in diesem Fall der Kundenabwanderung. Das Kundenverhalten im Prognosefenster wird über eine Abwanderungswahrscheinlichkeit im Intervall von 0 bis 1 definiert: die 1 steht hierbei für die sichere Abwanderung des Kunden, die 0 für das sichere Verbleiben des Kunden in der Geschäftsbeziehung.

Für die Zielsetzung der Abwanderungsprognose wird ein Prognosemodell gesucht, das den Zusammenhang zwischen Kundeneigenschaften im Kalibrierungsfenster und Kundenverhalten im Prognosefenster möglichst gut beschreibt.

Der Umfang von Kalibrierungs- und Prognosefenster sowie die Verwendung mehrerer Fenster im Rahmen von Modell-Training und -Evaluation werden im Abschnitt 5.4.1 zum Testdesign detaillierter erläutert.

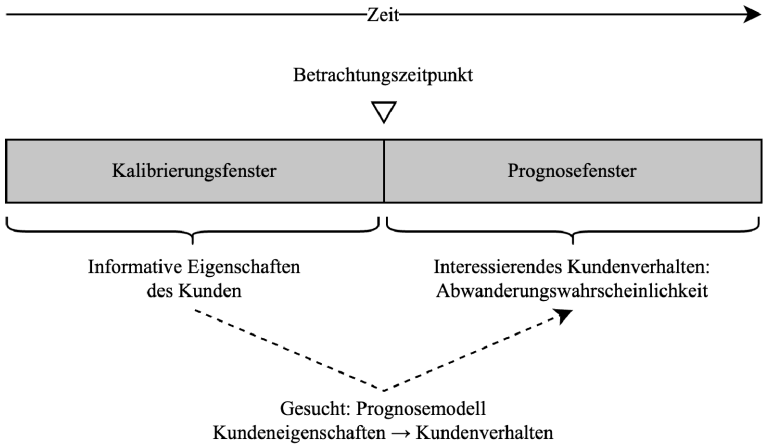

Abbildung 5: Formulierung der Abwanderungsprognose

Leistungsbewertung

Die Art der berücksichtigten Informationen bei der Untersuchung des Verhaltens sowie der Leistung eines Abwanderungsprognosemodells erlaubt die Unterscheidung von drei grundsätzlichen Arten von Verfahren zur Leistungsbewertung (vgl. Japkowicz, Shah 2011, S. 81):

- *Schwellenwertabhängige Verfahren* setzen eine deterministische Klassifikation voraus, d.h. sie legen für das betrachtete Modell fest, ab welchem Wahrscheinlichkeitsschwellenwert ein Kunde als Abwanderer klassifiziert wird. Auf Basis dieser Festlegung und einem Testdatensatz wird eine Konfusionsmatrix (vgl. Ting 2016) berechnet, die als Grundlage für verschiedene Metriken wie bspw. der Prognosegenauigkeit dient.

- *Schwellenwertunabhängige Verfahren* bewerten die Modell-Leistung allgemeiner auf Basis der prognostizierten Abwanderungswahrscheinlichkeiten. Sie sind unabhängig von einer bestimmten Entscheidungsschwelle. Häufig kommen hierbei graphische Verfahren zum Einsatz, die die Leistung eines Modells bei variierenden Schwellenwerten illustrieren. Ergänzend lassen sich die graphischen Verfahren zu eindimensionalen Metriken verdichten.

- *Kosten-Nutzen-basierte Verfahren* beziehen neben den prognostizierten Abwanderungswahrscheinlichkeiten weitere Informationen zum geplanten Einsatz des Modells in die Bewertung ein. So können bspw. Informationen zu erwarteten Kosten bzw. Erträgen von falschen bzw. korrekten Prognosen zur Berechnung von erwarteten Ergebnisbeiträgen von Kundenbindungsmaßnahmen dienen (vgl. Verbraken u. a. 2013).

Im Fall der WER GmbH sind die konkreten Maßnahmen zur Kundenbindung zum Zeitpunkt der Modellerstellung noch nicht ausgestaltet, so dass kein Kosten-Nutzen-basiertes Bewertungsverfahren in Frage kommt. Die Untersuchung des Verhaltens sowie der Leistung des untersuchten Modells basiert daher auf zwei im Kontext von Abwanderungsprognosen häufig verwendeten graphischen Analysemethoden und damit verbundenen Kennzahlen.

Die erste verwendete Methode ist die Grenzwertoptimierungskurve (engl.: receiver operating characteristic, ROC). Die ROC-Kurve stellt die Beziehung zwischen Sensitivität und Spezifität eines Klassifikationsmodells in Abhängigkeit von verschiedenen Klassifikationsschwellenwerten dar (vgl. Flach 2011, S. 869 ff.; Japkowicz, Shah 2011, S. 112 ff.). Als zusammenfassende, schwellenwertunabhängige Kennzahl für den Vergleich verschiedener ROC-Kurven dient die Fläche unter der ROC-Kurve (engl.: area under the ROC-curve, ROC-AUC). Sie liegt im Wertebereich zwischen 0 und 1 und kann als Wahrscheinlichkeit dafür interpretiert werden, dass das Modell eine zufällig ausgewählte positive Test-Instanz höher bewertet als eine negative (vgl. Flach 2011, S. 871; Japkowicz, Shah 2011, S. 129 ff.). Die Metrik beschreibt damit, wie akkurat das Klassifikationsmodell für verschiedene Schwellenwerte zwischen den beiden Klassen unterscheiden kann.

Die zweite verwendete Methode, die Lift-Kurve, ist in der betrieblichen Praxis weiter verbreitet. Sie stellt den Anteil an richtig positiven Prognosen im Verhältnis zum Anteil der gesamten Test-Datenmenge dar (vgl. Japkowicz, Shah 2011, S. 132). Als zusammenfassende Metrik eignet sich der Top-Decile-Lift (TDL). Die Kennzahl teilt die Präzision der am höchsten bewerteten 10% der Test-Instanzen durch die relative Häufigkeit der positiven Klasse. Sie drückt damit aus, wie viel besser ein Modell die Abwanderer im obersten Dezil im Vergleich zu einer zufälligen Selektion von Kunden identifiziert (vgl. Gattermann-Itschert, Thonemann 2022, S. 140).

Die genannten graphischen Verfahren und korrespondierenden Metriken erlauben allein für sich genommen noch keine sinnvolle Beurteilung der Modell-Leistung. Die Prognoseleistung des entwickelten Modells wird daher mit relevanten Vergleichsmaßstäben (*Baselines*) gegenübergestellt, um eine Leistungsverbesserung aufzeigen zu können bzw. den Mehrwert der analytischen Datennutzung zu quantifizieren (vgl. Provost, Fawcett 2013, S. 204 ff.).

Als analytisches Erfolgskriterium soll das entwickelte Modell die Abwanderungswahrscheinlichkeit von Kunden signifikant besser prognostizieren als die folgenden, zunehmend komplizierten Baselines:

1. Ein naives Modell, das die Abwanderungswahrscheinlichkeit eines Kunden zufällig „rät“. Ein Zufallsmodell zu übertreffen erscheint relativ leicht. Es kann jedoch bereits aufzeigen, dass für die Aufgabenstellung relevante Informationen aus den analysierten Daten abgeleitet werden können.

2. Eine Management-Heuristik, die eine einfache Entscheidungsregel basierend auf Domänen-Wissen und Erfahrung ableitet. Die „Pausen-Heuristik“, d.h. die Prognose der Kundenabwanderung anhand der Aktualität des letzten Kaufs ist in der Praxis weit verbreitet und funktioniert in vielen Fällen gleich gut wie komplexere statistische Modelle (vgl. Wübben, Wangenheim 2008, S. 84 ff.).

3. Ein aus dem Direktmarketing stammendes RFM-Verfahren (vgl. Gupta u. a. 2006, S. 142; Kumar, Reinartz 2018, S. 102 ff.), das Kunden in jeweils fünf gleich große Gruppen in Bezug auf Aktualität („R“), Frequenz („F“) und Wert („M“) ihres Kaufverhaltens klassifiziert und aus den entstehenden 5x5x5=125 Gruppen oder Zellen eine Bewertung für die Kundenbindung ableitet.

4. Ein einfaches, daten-reduziertes Klassifikationsmodell, das nur eine geringe Menge an Informationen aus den Daten berücksichtigt. Verwendet wird das Modell 1R (vgl. Holte 1993), das eine Klassifikation anhand der Ausprägungen genau eines Merkmals mit dem höchsten Informationsgehalt vornimmt. Das Modell entspricht einem 1-stufigen Entscheidungsbaum und wird häufig auch als „Entscheidungsstumpf“ (engl.: decision stump) bezeichnet (vgl. Iba, Langley 1992).

5. Eine logistische Regression als statistisches Verfahren. Die logistische Regression galt aufgrund der einfachen Implementierung, der angemessenen Leistung und der guten Interpretierbarkeit der Ergebnisse lange Zeit als „Gold-Standard“ für die Abwanderungsprognose (vgl. Bogaert, Delaere 2023, S. 4).

5.2 Datenverständnis

5.2.1 Daten sammeln

Die zentrale Datenquelle für das Projekt bildet das System CDH, ein Transaktionssystem, welches u.a. die operativen Verkaufs-, Auftragsabwicklungs- und Kundenbearbeitungsprozesse der WER GmbH für die Kunden im Streckengeschäft unterstützt. Aus dem System wurden die folgenden Datensammlungen in Form von tabellenorientierten Dateien für die weitere Verarbeitung extrahiert:

- *Kundenstammdaten* mit Angaben bspw. zu Adressen, Kundenherkunft (Organisch vs. Zukauf), Zahlungskonditionen und zugeordneten Vertriebsmitarbeitern.

- *Auftrags- bzw. Rechnungsdaten* auf Kopf- und Positionsebene mit Angaben bspw. zu verkauften Artikeln, Mengen, Preisen und Rabatten.

- *Kundenkontaktdaten* mit einer Historie von Kundenkontakten und Angaben bspw. zu Kommunikationspartnern und -kanälen.

Die genannten Datensammlungen bilden die Ausgangsbasis für die spätere Datenintegration und die Bildung von Merkmalen im Rahmen der Datenaufbereitung (vgl. Abschnitt 5.3).

5.2.2 Daten erkunden

Die Datenexploration beschreibt die grundlegenden Eigenschaften der gesammelten Daten und leitet allgemeine Erkenntnisse ab, die für den weiteren Analyseprozess hilfreich sein können (vgl. Berthold u. a. 2020, S. 33; Chapman u. a. 2010, S. 18). Im Kontext der Abwanderungsprognose sind insbesondere Erkenntnisse bezogen auf das Kaufverhalten der Kunden von Interesse.

Abbildung 6 illustriert, dass das Unternehmen die Anzahl der aktiven Kunden im betrachteten Geschäftsbereich und Zeitraum sowohl über organisches Wachstum als auch über den Aufkauf von Wettbewerbern auf zuletzt 967 aktive Kunden im Jahr 2023 steigern konnte.

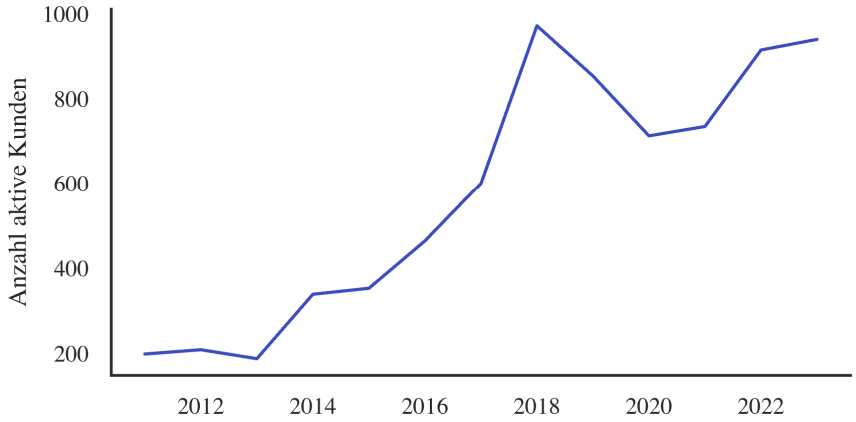

Abbildung 6: Entwicklung aktiver Kunden

Abbildung 7 verdeutlicht die Potenziale im Bereich der Kundenbindung. Im Schnitt bestehen die aktiven Kunden eines Jahres zu 59% aus Bestandskunden. Diese tragen mit einem durchschnittlichen Umsatzanteil von 78% im Vergleich zum Neukundengeschäft überproportional zum Umsatz bei. Aus einer anderen Perspektive betrachtet „verliert“ das Unternehmen jedes Jahr einen signifikanten Anteil seiner aktiven und potenziell ertragreichen Kunden und muss das verlorene Geschäft über Neukunden kompensieren.

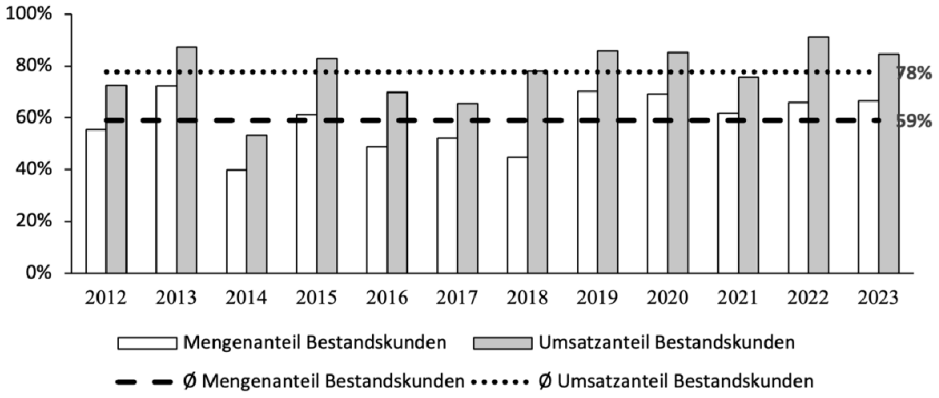

Abbildung 7: Mengen- und Umsatzanteile Bestandskunden

Im Mittel über alle Jahre des Betrachtungszeitraums realisierten die gesamten aktiven Kunden einen mittleren Jahresumsatz von ca. 8‘460 € je Kunde, wobei die Umsätze relativ stark variierten (vgl. Abbildung 8).

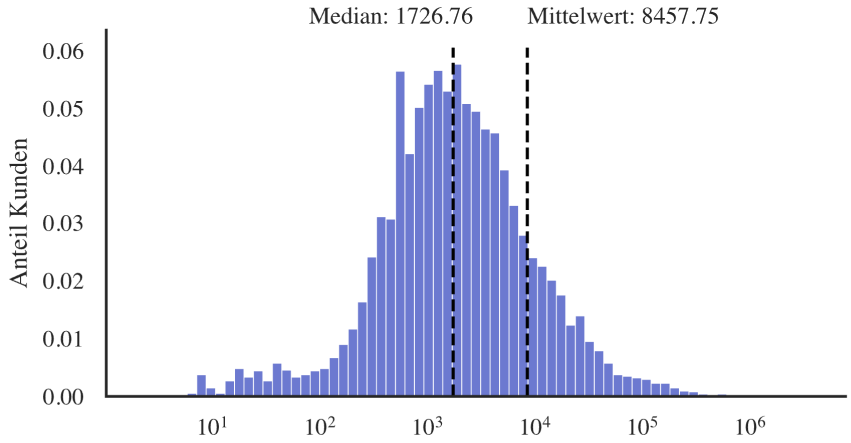

Abbildung 8: Verteilung der Jahresumsätze

Analog zu den Umsätzen unterschieden sich auch die Kauffrequenzen und Wiederkaufzeiten zwischen den Kunden vergleichsweise stark (vgl. Abbildung 9 und Abbildung 10). Während die Hälfte der Kunden weniger als vier Aufträge pro Jahr platzierten, existieren auch wenige Kunden mit vierstelligen Auftragszahlen pro Jahr. Die mittlere Wiederkaufszeit pro Kunde liegt bei knapp 262 Tagen, wobei knapp 78% der Kunden eine mittlere Wiederkaufszeit von weniger als 365 Tagen bzw. einem Jahr haben.

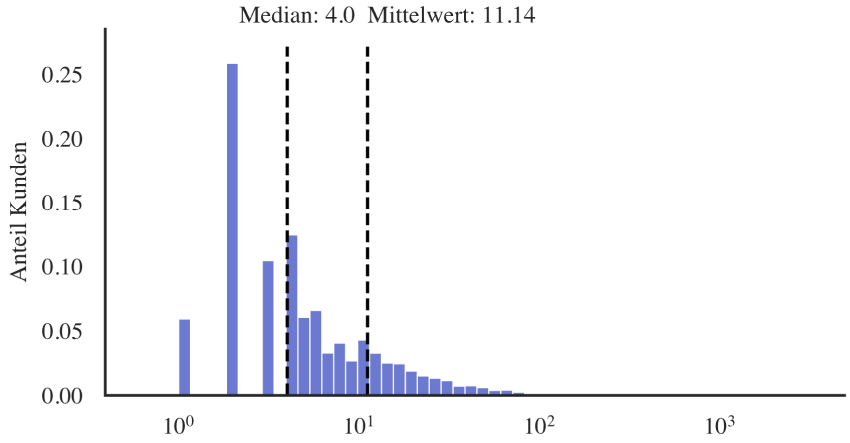

Abbildung 9: Verteilung der Kauffrequenzen

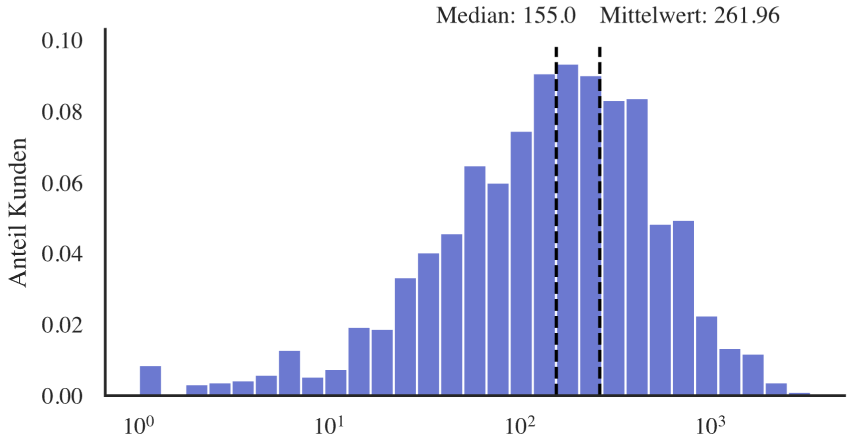

Abbildung 10: Verteilung der Wiederkaufzeiten

5.3 Datenaufbereitung

5.3.1 Daten auswählen & integrieren

Den Ausgangspunkt der Datenaufbereitung bildet die Auswahl einer Menge von Datensätzen und Attributen aus den gesammelten Daten (vgl. Abschnitt 5.2.1), sowie die Integration der einzelnen Daten auf die Ebene des individuellen Kunden. Beispielhafte Aufgaben in dieser Phase betreffen die zeitliche Selektion der Daten, die Eingrenzung auf Rechnungsbelege, die Ausgrenzung von irrelevanten Artikelgruppen sowie die Aggregation von Kunden mit mehreren lokalen Abnehmern bzw. Lieferadressen. Die resultierende Datensammlung bildet den Ausgangspunkt für die folgende Konstruktion von Merkmalen.

5.3.2 Merkmale konstruieren

Die Konstruktion von Merkmalen nutzt Domänenwissen um die historischen Daten in Darstellungsformen zu überführen, die die Effektivität der zu bildenden Modelle bzw. die erwartete Modellleistung erhöhen (vgl. Berthold u. a. 2020, S. 140; Kuhn, Johnson 2020, S. 3).

Merkmale zur Prognose der Kundenabwanderung lassen sich anhand der primär betrachteten Datenobjekte in fünf inhaltliche Kategorien gruppieren (vgl. Naß 2012, S. 75 ff.):

- *Transaktionsbezogene Merkmale* fassen Informationen zu Transaktionen zwischen Unternehmen und Kunde zusammen.
- *Leistungs- bzw. angebotsbezogene Merkmale* beziehen sich auf die Unternehmensleistung und umfassen bspw. Informationen zu Preisen und Konditionen.
- *Kundenbezogene Merkmale* umfassen beschreibende Charakteristika des Kunden.
- *Interaktionsbezogene Merkmale* umfassen Kontaktinformationen zwischen Unternehmen und Kunde, bspw. in Form von kundeninitiierter Kommunikation oder Marketingkommunikation.
- *Wettbewerbsorientierte Merkmale* fassen Informationen zu externen Aspekten der Kundenbeziehung zusammen, die diese festigen oder gefährden können.

Darüber hinaus lassen sich Merkmale anhand ihres Zeitbezugs unterscheiden. Zum Einen betrifft die Unterscheidung die Veränderung der Merkmale im Zeitverlauf (vgl. Tecklenburg 2008, S. 185):

- Im Zeitverlauf *konstante (zeitpunktunabhängige) Merkmale* beschreiben genau eine Datenausprägung im gesamten Betrachtungszeitraum. Beispiele sind der geographische Standort eines Kunden oder seine gesamte Beziehungsdauer.

- Im Zeitverlauf *variierende (zeitpunktabhängige) Merkmale* können mehrfach im Betrachtungszeitraum - bspw. je Transaktion oder je Teilzeitraum - beobachtet werden. Beispiele sind Umsätze je Transaktion oder Zwischenkaufzeiten.

Zum anderen können Merkmale dahingehend unterschieden werden, auf Basis welchen Betrachtungszeitraums sie berechnet werden (vgl. Mirkovic u. a. 2022, S. 7):

- *Kumulative Merkmale* beziehen sich auf alle Daten zwischen dem Datum des ersten verfügbaren Datensatzes und dem Prognosezeitpunkt.

- *Delta-Merkmale* hingegen beziehen sich auf einen klar abgrenzten Zeitraum vor dem Prognosezeitpunkt (vgl. Abschnitt 5.4.1).

Variierende Merkmale werden in der Datenaufbereitung über passende Aggregationen auf einen Wert je Betrachtungszeitraum verdichtet. Abhängig vom Merkmal kommen hierbei verschiedene Kennzahlen zum Einsatz:

- *Lagekennzahlen* wie Minimum, Maximum, Mittelwert oder Median charakterisieren die zentrale Lage der variierenden Merkmale.

- *Streuungskennzahlen* wie Varianz oder Variationskoeffizient beschreiben die Schwankung der Merkmale.

- *Entwicklungskennzahlen* wie die Steigung einer linearen Regression (Slope) oder das Verhältnis des letzten Werts eines Merkmals zu seinem Mittelwert bilden Informationen über die Veränderung bzw. Veränderungsrichtung der Merkmale ab.

Die folgenden Abschnitte motivieren die Konstruktion einzelner Merkmale in den verschiedenen Kategorien. Tabelle 3 im Anhang fasst die insgesamt 76 gebildeten Merkmale übersichtsartig zusammen.

Transaktionsbezogene Merkmale

Transaktionsbezogene Merkmale des Kundenverhaltens nehmen häufig eine herausgehobene Stellung ein, da diese relativ einfach zu messen sind, häufig auf Basis von Transaktionsdaten in Unternehmen vorliegen und gute Prädiktoren für das zukünftige Kaufverhalten darstellen (vgl. Schmittlein, Peterson 1994, S. 45). Sie lassen sich feiner in drei Unterkategorien anhand des RFM-Akronyms unterteilen:

- **R**ecency-Merkmale messen die Aktualität der letzten Transaktion.

- **F**requency-Merkmale beschreiben die Häufigkeit von Transaktionen.

- **M**onetary Value-Merkmale fassen wertbezogene Eigenschaften von Transaktionen zusammen.

Die Dauer seit der letzten Transaktion (Recency) kann im nicht-vertraglichen Kontext ein wichtiger Indikator für Aktivität oder Inaktivität der Kundenbeziehung sein: Kunden die erst kürzlich eingekauft haben sind eher aktiv als Kunden, deren letzter Einkauf lange zurück liegt (vgl. Buckinx, Van den Poel 2005, S. 259).

Recency allein ist jedoch häufig nicht aussagekräftig, da Kunden ggf. in unterschiedlichen Frequenzen einkaufen. Das Merkmal wird daher ergänzt über Merkmale zur Transaktionshäufigkeit, die häufig positiv mit zukünftigen Transaktionen des Kunden korreliert (vgl. Buckinx, Van den Poel 2005, S. 260). Berechnet werden sowohl die Anzahl an Transaktionen pro Betrachtungszeitraum (Frequency) als auch die Zeiten zwischen einzelnen Käufen (Zwischenkaufzeiten). Die Frequency wird neben dem gesamten Betrachtungszeitraum für den letzten Monat, das letzte Quartal sowie das letzte Jahr berechnet, da jüngere Zeiträume wichtiger für die Prognose sein könnten (vgl. Buckinx, Van den Poel 2005, S. 260). Ergänzend wird die Anzahl an Transaktionen in Relation zur Beziehungsdauer in Tagen gesetzt.

Verschiedene Aggregationen der Zwischenkaufzeiten können Indikationen zur Kundenbindung liefern. So spiegelt die mittlere Zwischenkaufzeit die Recency über den gesamten Betrachtungszeitraum wieder während die Streuung der Zwischenkaufzeiten einen weiteren Indikator der Kundenbindung darstellt: unregelmäßige Einkäufe können auf weniger loyale Kunden hinweisen (vgl. Buckinx, Van den Poel 2005, S. 260). Eine Verkürzung der Zwischenkaufzeiten im Zeitablauf kann auf eine Zunahme der Kundenbindung hinweisen. Das Verhältnis von Recency und mittlerer Zwischenkaufzeit kann darüber hinaus einen genaueren Hinweis zur Kundenbindung liefern: eine Relation größer als eins deutet in diesem Fall ggf. auf ein Kundenbindungsproblem hin (vgl. Ascarza u. a. 2018, S. 67).

Neben Aktualität und Häufigkeit von Transaktionen ist der monetäre Wert des bisherigen Kaufverhaltens eines jeden Kunden in der Regel gut geeignet, das zukünftige Kaufverhalten vorherzusagen (vgl. Buckinx, Van den Poel 2005, S. 260; Schmittlein, Peterson 1994, S. 43). Betrachtet werden zum einen der Umsatz je Transaktion und zum anderen der monatliche Umsatz der zurückliegenden 12 Monate. Für diese beiden variierenden monetären Merkmale werden verschiedene, in Tabelle 3 aufgelistete Lage-, Streuungs- und Entwicklungskennzahlen berechnet. Zusätzlich wird der Gesamtumsatz des Kunden sowie der Umsatz im letzten Monat, Quartal und Jahr berechnet. Ergänzt werden die Umsatzkennzahlen über die Relation zwischen Gesamtumsatz und Beziehungsdauer in Tagen.

Neben dem Umsatz wird die Unternehmensmarge als Einflussgröße auf die Kundenbindung betrachtet (vgl. Gattermann-Itschert, Thonemann 2022, S. 137 f.). Berechnet wird die Differenz aus Umsatz und Kosten je Transaktion monatsweise für die letzten 12 Monate. Die Aggregationen entprechen den für den Umsatz gebildeten Kennzahlen.

Leistungs- bzw. Angebotsbezogene Merkmale

Das Kaufverhalten über die Angebotsbreite kann ein Indikator für Kundenbindung sein (vgl. Buckinx, Van den Poel 2005, S. 260). Eine Hypothese ist bspw. das Kunden zunächst mit dem Kauf einer Produktkategorie einsteigen und im für das Unternehmen positiven Fall ihren Umsatz auf weitere Teile des Sortiments ausweiten. Als Merkmal wird der relative Anteil des Gesamtumsatzes eines Kunden je Produktkategorie (5 insgesamt) berechnet. Zusätzlich wird die Anzahl der jemals gekauften Produktkategorien ermittelt.

Kundenbezogene Merkmale

Als kundenbezogenes Merkmal wird der Ort des Kunden einbezogen. Berechnet werden die ersten beiden Ziffern der Postleitzahl in Deutschland bzw. das Länderkürzel im Fall internationaler Kunden.

Die WER GmbH hat im Lauf ihrer Geschichte mehrere Werbemittelhändler übernommen. Für die Analyse wird unterschieden, ob ein Kunde organisch vom Unternehmen akquiriert wurde oder die Kundenbeziehung über einen Aufkauf initiiert wurde.

Interaktionsbezogene Merkmale

Die Beziehungsdauer zwischen Kunde und Unternehmen zeigt häufig einen positiven Zusammenhang mit der zukünftigen Stabilität der Geschäftsbeziehung sowie mit der Breite der Angebotsnutzung (vgl. Buckinx, Van den Poel 2005, S. 261). Gemessen wird in dieser Arbeit die Dauer zwischen dem aktuellen Betrachtungszeitpunkt und dem ersten Kauf des Kunden in Tagen. Darüber hinaus können Kontakte zwischen Unternehmen und Kunde die Beziehung stärken. Als Merkmale werden die Gesamtanzahl der Kundenkontakte, die Zeit seit dem letzten Kontakt in Tagen sowie die Art des letzten Kontakts (Telefonat, Email, etc.) ermittelt. Der dem Kunden zugeordnete Vertriebsmitarbeiter wird ebenfalls in die Analyse einbezogen.

Wettbewerbsorientierte Merkmale

Die Analyse bezieht keine wettbewerbsorientierten Merkmale ein.

5.3.3 Merkmale auswählen

Die Merkmalsauswahl untersucht die Relevanz der berechneten Merkmale für die Lösung des Prognoseproblems. Die Zielsetzung besteht darin, die Anzahl verwendeter Prädiktoren in der Modellbildung zu reduzieren ohne die Prognoseleistung zu beeinträchtigen (vgl. Berthold u. a. 2020, S. 127 f. ; Kuhn, Johnson 2020, S. 227 f.).

Ein geeigneter Ausgangspunkt zur Untersuchung von Merkmalen liegt in der Analyse des bivariaten Zusammenhangs zwischen den einzelnen Merkmalen und der Zielvariablen. Eine geeignete Größe zur Quantifizierung dieses Zusammenhangs ist der *Informationsgewinn* bzw. die *Transinformation* (engl.: *mutual information*) (vgl. Brownlee 2019; Murphy 2022, S. 211 ff.). Es handelt sich dabei um einen nichtnegativen Wert, der die Abhängigkeit zwischen zwei Variablen quantifiziert. Im Anwendungsgebiet der Merkmalsauswahl beschreibt der Informationsgewinn die Menge an Informationen, die man über die Zielvariable durch Beobachtung des betrachteten Merkmals erhält. Er ist gleich

Null, wenn Merkmal und Zielvariable komplett unabhängig voneinander sind und bildet größere Abhängigkeiten über höhere Werte ab.

Abbildung 11 illustriert die Ausprägungen für die 10 Merkmale mit dem höchsten sowie die 10 Merkmale mit dem niedrigsten Informationsgewinn. Die Abbildung verdeutlicht, dass die transaktionsbezogenen, monetären Merkmale wie Umsatz und Marge in verschiedenen Aggregationen isoliert betrachtet den höchsten Erklärungsbeitrag zur Unterscheidung von abwandernden und treuen Kunden liefern. Kunden- und beziehungsbezogene Merkmale hingegen sind deutlich weniger als alleinstehende Prädiktoren geeignet.

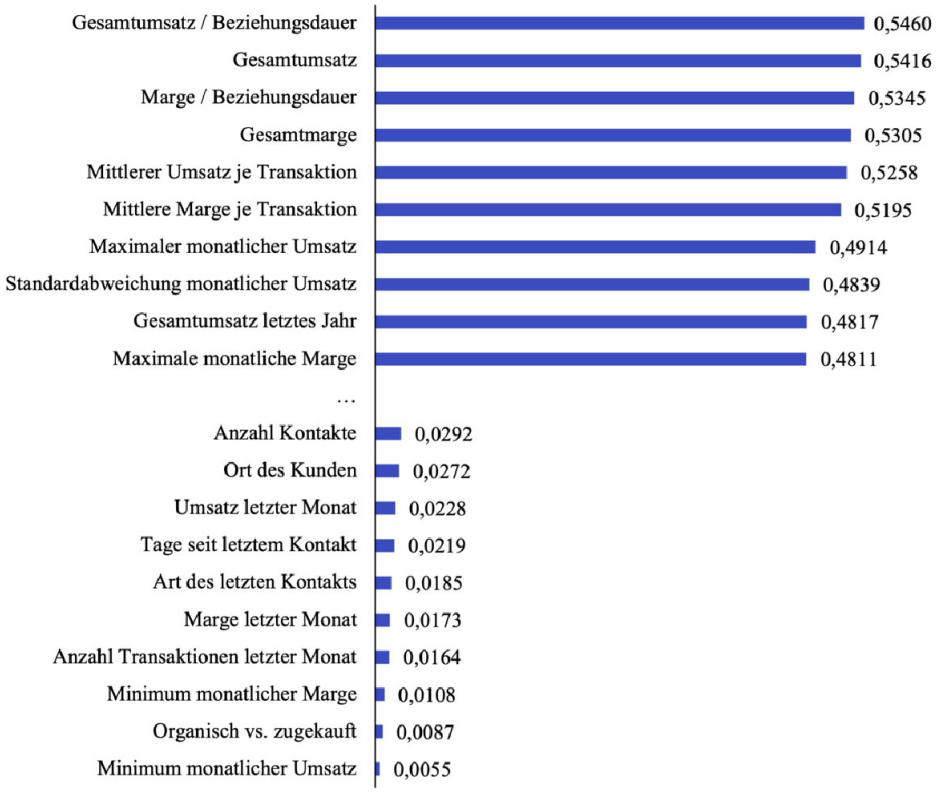

Abbildung 11: Merkmale mit dem höchsten und niedrigsten Informationsgewinn

Abbildung 12 und Abbildung 13 illustrieren beispielhaft die Eignung von umsatzbezogenen Kennzahlen als Prädiktoren für die Abwanderungsprognose. Sowohl der kumulierte Gesamtumsatz zum Betrachtungszeitpunkt als auch der Gesamtumsatz in Relation zur Beziehungsdauer fallen für Abwanderer im Mittel geringer aus als für wiederkaufende Kunden.

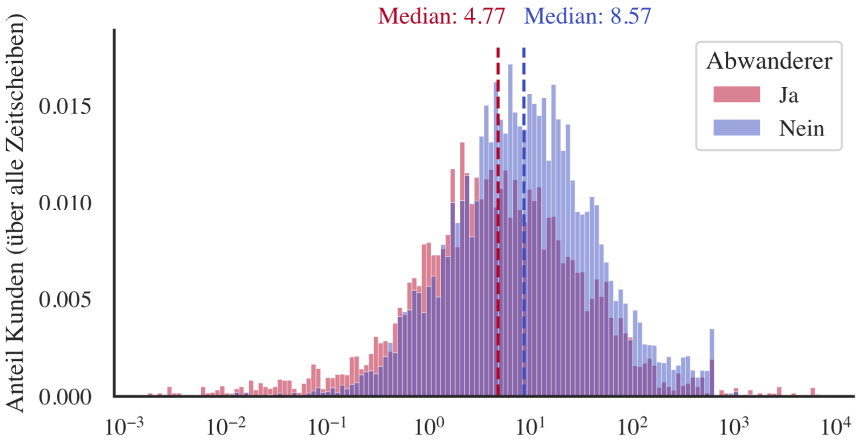

Abbildung 12: Verteilung des Merkmals Gesamtumsatz / Beziehungsdauer unter Abwanderern und Nicht-Abwanderern

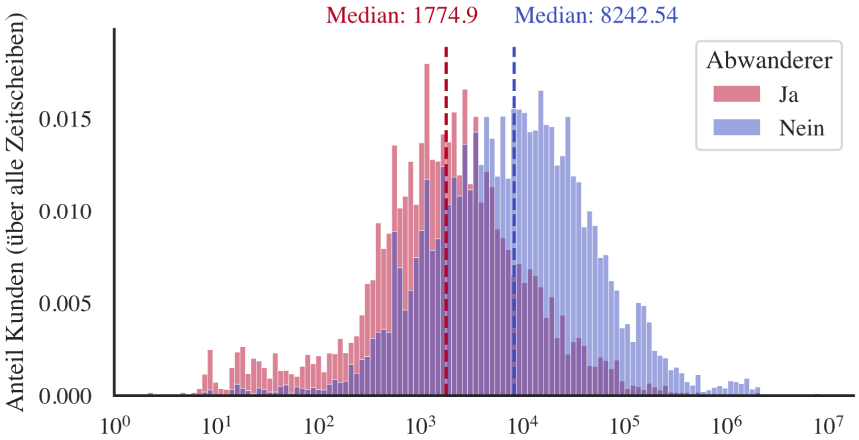

Abbildung 13: Verteilung des Merkmals Gesamtumsatz unter Abwanderern und Nicht-Abwanderern

Verfahren zur systematischen Auswahl von Merkmalen lassen sich in drei Kategorien unterteilen (vgl. Kuhn, Johnson 2020, S. 228 ff.):

- *Intrinsische Verfahren* sind im jeweiligen Trainingsalgorithmus integriert. Beispiele sind die Auswahl von Prädiktoren und Schnittpunkten zur Aufteilung von Datenmengen in baumbasierten Verfahren oder Regularisierungsansätze wie die Lasso- oder Ridge-Regression.

- *Filterverfahren* analysieren zunächst die einzelnen Prädiktoren auf deren Relevanz und geben im Anschluss nur ausgewählte an die Modellbildung weiter. Bspw. könnten Merkmale anhand der Höhe ihres Informationsgewinns ausgewählt oder ausgeschlossen werden.

- *Wrapper-Methoden* verwenden iterative Suchverfahren, bei denen dem Modell wiederholt Teilmengen von Prädiktoren zugeführt werden und die sich daraus ergebende Schätzung der Modellleistung als Richtschnur für die Auswahl der Merkmale dient. Ein Beispiel ist die rekursive Merkmalseliminierung.

Der Großteil der im folgenden Abschnitt 5.4.2 ausgewählten Algorithmen verfügt über integrierte intrinsische Verfahren zur Merkmalsauswahl. Diese sind i.d.R. deutlich effizienter in der Ausführung als Wrapper-Verfahren und bieten im Vergleich mit den intrinsischen Methoden eine direkte Verbindung zur Zielfunktion der Algorithmen. Aus diesen Gründen verzichtet die vorliegende Analyse auf eine weitergehende, separate Merkmalsauswahl über Filter- oder Wrapper-Verfahren.

5.4 Modellierung

5.4.1 Testdesign festlegen

Das Testdesign dient der Überprüfung der Modellqualität (vgl. Chapman u. a. 2010, S. 24) und verfolgt dabei drei grundlegende Ziele (vgl. Raschka 2020, S. 4):

1. Die Abschätzung der Verallgemeinerungsleistung des Modells, d.h. der Prognoseleistung des Modells auf zukünftigen (unbekannten) Daten.

2. Die Optimierung der Prognoseleistung des Modells durch Anpassung der Eigenschaften des Lernalgorithmus.

3. Den Vergleich verschiedener (optimierter) Modelle untereinander und die Identifikation des für die Problemstellung besten geeigneten Modells.

Es unterteilt die Daten hierzu mehrfach in Training-, Test- und Validierungsmengen (vgl. Wikipedia 2024):

- Trainingsdaten dienen der Anpassung des Modells an die Daten.

- Validierungsdaten dienen der optimalen Einstellung des Lernalgorithmus bspw. über Merkmalsselektion oder Hyperparameter-Optimierung.

- Testdaten liefern eine unvoreingenommene Bewertung der Leistung eines optimierten Modells.

Der Aufbau des Testdesigns folgt einer zeitbasierten, verschachtelten Kreuzvalidierung (vgl. Hyndman, Athanasopoulos 2021; Provost, Fawcett 2013, S. 135; Raschka 2020, S. 45 f.). Abbildung 14 stellt die Unterteilung der Daten im Überblick dar.

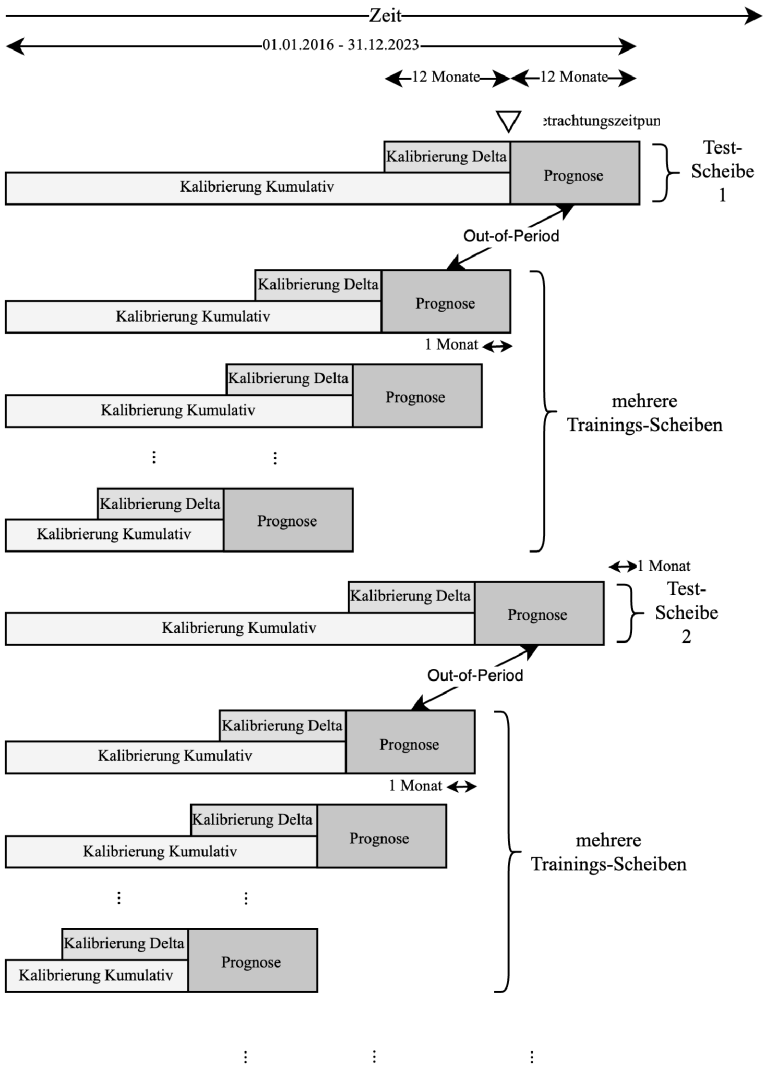

Abbildung 14: Überblick Test-Design (in Anlehnung an Gattermann-Itschert, Thonemann 2022, S. 138; Mirkovic u. a. 2022, S. 8)

Eine Testdatenmenge bezieht sich auf einen konkreten Betrachtungszeitpunkt und umfasst jeweils einen Zeitraum von 12 Monaten als Prognosefenster sowie sämtliche verfügbaren Daten davor als Kalibrierungsfenster (vgl. Abschnitt 5.1.2). Die aktuellsten 12 Monate des Kalibrierungsfensters dienen der Berechnung von Delta-Merkmalen während das gesamte Kalibrierungsfenster die Berechnung von kumulativen Merkmalen ermöglicht (vgl. Abschnitt 5.3.2).

Die Trainingsdatenmenge endet jeweils 12 Monate vor Beginn der zugeordneten Testdatenmenge. Hierdurch basiert die Leistungsbewertung des Modells nur auf Daten, die zum Trainingszeitpunkt noch nicht vorhanden waren. Dieser sog. „Out-of-Period"-Ansatz berücksichtigt die inhärente Zeitverschiebung zwischen der Erstellung eines Modells sowie dessen Anwendung in der Praxis und liefert somit eine realistischere Einschätzung der Prognoseleistung als eine „Out-of-Sample"-Bewertung (vgl. Gattermann-Itschert, Thonemann 2022, S. 137).

Das Testdesign verwendet für jede Testdatenmenge mehrere Betrachtungszeitpunkte bzw. Trainingsscheiben (vgl. Gattermann-Itschert, Thonemann 2022, S. 137; Gür Ali, Arıtürk 2014, S. 7892 ff.; Mirkovic u. a. 2022, S. 7 f.). Die Betrachtungszeitpunkte der Trainingsdatenmengen werden dabei jeweils um einen Monat in die Vergangenheit verschoben. Dieser in der Literatur als „multi-slicing" oder "Multi Period Training" bezeichnete Ansatz verspricht eine verbesserte Modell-Leistung durch eine Vergrößerung der zur Verfügung stehenden Trainingsdatenmenge sowie eine Berücksichtigung verschiedener Betrachtungszeitpunkte beim Training (vgl. Gattermann-Itschert, Thonemann 2021, S. 673).

Das Testdesign kombiniert mehrere Testdatenmengen in einer verschachtelten Kreuzvalidierung: während die Modelle in einer inneren Schleife optimiert werden, bewertet eine äußere Schleife die Modell-Leistung (vgl. Provost, Fawcett 2013, S. 135; Raschka 2020, S. 45 f.). Die äußere Schleife bildet 12 Kombinationen aus Test- und Trainingsdatenmengen, wobei der Betrachtungszeitpunkt beginnend mit dem aktuellst möglichen jeweils um einen Monat in die Vergangenheit verschoben wird. Die innere Schleife optimiert Hyperparameter der verschiedenen Lernverfahren basierend auf einer 6-fachen Kreuzvalidierung analog zur äußeren Schleife. Als Optimierung kommt hierbei ein Verfahren der Bayes'schen Optimierung in Form des *Tree Parzen Estimators* zum Einsatz (Bergstra u. a. 2011, S. 2550 f.; Feurer, Hutter 2019, S. 9 ff.).

Die Anzahl verwendeter Trainingsscheiben wird empirisch über eine Analyse von Lernkurven (vgl. Perlich 2011, S. 577 ff.) ermittelt. Die Ergebnisse in Abbildung 15 zeigen dabei die Generalisierungsleistung dreier verschiedener Modelle in Form der ROC-AUC auf mehreren Test-Sets in Abhängigkeit der Anzahl verwendeter Trainingszeitscheiben.

Die Kurven wurden analog zum bisher skizzierten Test-Design in einer 12-fachen, zeitbasierten Kreuzvalidierung berechnet, wobei die Modelle ohne Optimierung mit Standard-Hyperparametern trainiert wurden. Als Modelle dienten ein statistisches Verfahren in Form der logistischen Regression (LR) (vgl. Kalisch, Meier 2021), ein Bagging-basiertes homogenes Ensemble in Form der Extremely Randomized Trees (XT) (vgl.

Geurts u. a. 2006) sowie ein Boosting-basiertes homogenes Ensemble in Form eines Gradient Boosting Classifiers (GB) (vgl. Friedman 2001).

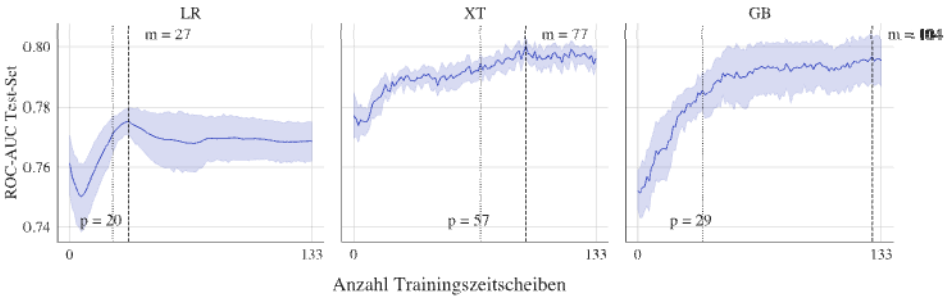

Abbildung 15: Lernkurvenanalyse

Die Ergebnisse veranschaulichen, dass alle drei Modelle vom steigenden Umfang der Trainingsdaten profitieren, wobei die Steigerung der Modell-Leistung im Verlauf abnimmt und sich auf einem Plateau stabilisiert. Gekennzeichnet sind je Modell die Anzahl an Zeitscheiben *m* mit der im Mittel besten Modell-Leistung sowie die Anzahl an Zeitscheiben *p*, bei denen die mittlere Modellleistung innerhalb einer Standardabweichung vom Maximum liegt. Die Ensemble-Verfahren profitieren deutlicher als das lineare Modell vom steigenden Umfang an Trainingsdaten und erreichen auch deutlich höhere ROC-AUC-Werte.

In Abwägung zwischen Modell-Leistung und Trainingslaufzeit erscheint eine Menge von 57 Test-Zeitscheiben gut geeignet. Diese Größe wird einheitlich im Rahmen des Testdesigns verwendet.

Zwecks besserer Vergleichbarkeit der Algorithmen erhält jedes Verfahren ein begrenztes Zeitbudget für das Training: je Testdatenmenge wird die Trainingszeit auf 60 min. beschränkt. Alle Algorithmen werden dabei auf der gleichen Hardware-Basis ausgeführt.

5.4.2 Algorithmen auswählen

Im Anschluss an das Testdesign sind konkrete Modellierungstechniken für die anschließende Bewertung und den Vergleich mit den beschriebenen Baseline-Verfahren (vgl. Abschnitt 5.1.2) auszuwählen. Die in Tabelle 2 zusammengefasste Auswahl konzentriert sich auf alleinstehende ML-Verfahren und ML-Ensembles, die auf der Basis bestehender Untersuchungen eine gute Prognoseleistung für den Anwendungsfall erwarten lassen.

Die Auswahl der alleinstehenden ML-Verfahren sowie der homogenen ML-Ensembles basiert auf dem dokumentierten Einsatz dieser Verfahren im Anwendungsgebiet der Kundenabwanderungsprognose (vgl. Bogaert, Delaere 2023, S. 1136 ff.) sowie ihrer generellen Eignung für Klassifikationsaufgaben.

Für den Vergleich mit heterogenen ML-Ensembles wurden bestehende Implementierung von *Automated Machine Learning* (AutoML) ausgewählt. AutoML-Implementierungen zielen darauf ab, verschiedene Aufgaben bzw. Phasen des Modellbildungsprozesses zu

automatisieren (vgl. Karmaker u. a. 2022, S. 2 ff.). Die drei ausgewählten Verfahren automatisieren im Schwerpunkt die Hyperparameteroptimierung sowie die Kombination verschiedener alleinstehender Verfahren zu heterogenen Ensembles und lieferten in Benchmarks auf verschiedenen Datensätzen gute Ergebnisse (vgl. Gijsbers u. a. 2024).

Ausgeschlossen vom Vergleich wurden (tiefe) neuronale Netze. Für den vorliegenden Anwendungsfall der Modellbildung auf tabellarischen Daten verspricht dieser Modelltyp i.d.R. keine Leistungsverbesserung im Vergleich zu ML-Ensembles wie bspw. Boosting-Verfahren und hat darüber hinaus den Nachteil signifikant längerer Trainingszeiten (vgl. Borisov u. a. 2024, S. 7513; Grinsztajn u. a. 2022, S. 511 f.)

Tabelle 2: Überblick ausgewählter Algorithmen

Klasse	**Algorithmus**	**Abkürzung**	**Beschreibung**
Alleinstehend	Decision Tree Classifier	DT	Entscheidungsbäume unterteilen die Trainingsinstanzen anhand von Regeln rekursiv in immer homogenere Teilmengen und Nutzen die entstehende Regelbasis zur Klassifikation (vgl. Friedman u. a. 2009, S. 305 ff.) .
	K-Nearest-Neighbour	KNN	Nächste-Nachbarn-Methoden klassifizieren eine Instanz auf der Grundlage der k ähnlichsten oder nächstgelegenen Instanzen (vgl. Friedman u. a. 2009, S. 463 ff.). Die Ähnlichkeit zwischen Instanzen wird anhand eines geeigneten Distanzmaßes ermittelt.
	Support Vector Machine	SVM	SVM verwendet den „Kernel-Trick“, um die Eingabe auf einen hochdimensionalen Merkmalsraum abzubilden (vgl. Géron 2022). Der Algorithmus konstruiert dann eine Hyperebene, die den Abstand zwischen den Ausgabeklassen maximiert.
Homogenes Ensemble	Random Forest	RF	RF kombiniert Bagging und eine Auswahl an Merkmalen mit Entscheidungsbäumen als Basismodell (vgl. Breiman 2001b).
	Extremely Randomized Trees	XT	XT funktioniert im Grundsatz analog zu RF, allerdings erfolgt die Auswahl an Merkmalen zufällig und nicht kriterienbasiert (vgl. Geurts u. a. 2006).

	XGBoost	XGB	XGB passt iterativ Entscheidungsbäume auf den Restfehler des Vorgängermodells zur Optimierung einer bestimmten Verlustfunktion an (vgl. Chen, Guestrin 2016). Das Boosting-Verfahren ist im Hinblick auf Konvergenz optimiert.
	LightGBM	LGBM	LGBM ist eine alternative Boosting-Implementierung, die Entscheidungsbäume blattweise aufbaut und eine histogrammbasierte Suche verwendet (vgl. Ke u. a. 2017).
	CatBoost	CAT	CAT ist eine weitere Boosting-Alternative, die im Hinblick auf die direkte Verwendung von kategorischen Merkmalen entwickelt wurde (vgl. Prokhorenkova u. a. 2018).
Heterogenes Ensemble	AutoSklearn	ASKL	AutoSklearn optimiert die Hyperparameter verschiedener alleinstehender Modelle und ML-Ensembles unter Verwendung einer Bayesschen Optimierung und kombiniert die Modelle im Anschluss zu einem gewichteten heterogenen Ensemble (vgl. Feurer u. a. 2015).
	Fast Library for Automated Machine Learning & Tuning	FAML	FAML optimiert eine Menge von alleinstehenden ML-Verfahren und homogenen Ensembles über ein mehrstufiges, randomisiertes und gerichtetes Suchverfahren (vgl. Wang u. a. 2021). Die optimierten Modelle werden im Anschluss über ein Stacking-Verfahren aggregiert.
	Tree-based Pipeline Optimization Tool	TPOT	TPOT stellt die Kombination verschiedener ML-Verfahren in baumartigen Strukturen dar, wobei die verschiedenen Zweige unterschiedliche Datenverarbeitungsschritte darstellen. Die Baumstrukturen werden über evolutionäre Algorithmen optimiert (vgl. Olson u. a. 2016).

5.4.3 Daten kodieren und transformieren

Abhängig von ihrem Skalenniveau müssen die erstellten Merkmale vor ihrer Eingabe in die verschiedenen Algorithmen ggf. über Kodierungen oder Transformationen aufbereitet werden. Diese Aufbereitung der Daten orientiert sich an den verschiedenen Algorithmen.

Für die linearen oder distanzbasierten Modelltypen werden ordinal oder nominal skalierte Eingaben über Dummy-Variablen abgebildet bzw. „One-Hot"-kodiert (vgl. Kuhn, Johnson 2020, S. 94 ff.). Darüber hinaus werden metrische Variablen z-standardisiert bzw. im Fall von KNN auf das Intervall [0,1] transformiert.

Für die baumbasierten Modelltypen werden ordinal oder nominal skalierte Variablen Label-kodiert (vgl. Kuhn, Johnson 2020, S. 114 ff.). Die metrischen Variablen werden für diese Algorithmen nicht transformiert.

Die AutoML-Implementierung der heterogenen ML-Ensembles umfassen eine automatisierte Aufbereitung der Eingabedaten, so dass die Merkmale in diesen Fällen nicht vorab aufbereitet werden.

5.4.4 Modelle erstellen und bewerten

Das in Abschnitt 5.4.1 beschriebene Testdesign liefert für jeden Algorithmus eine Bewertung der Modellleistung unter optimierten Hyperparametern für 12 verschiedene Kombinationen aus Trainings- und Test-Datenmengen. Abbildung 16 und Abbildung 17 illustrieren die Ergebnisse für die beiden Evaluationsmetriken ROC-AUC und TDL. Die Modelle sind in den Abbildungen nach ihrem mittleren Ergebnis über die verschiedenen Testdatenmengen aufsteigend von unten nach oben sortiert. Tabelle 4 im Anhang listet ergänzend die Mittelwerte und Standardabweichungen der einzelnen Modelle auf.

Mit Ausnahme der logistischen Regression liefern die Baseline-Verfahren im Mittel die schlechtesten ROC-AUC-Werte und weisen darüber hinaus eine relativ große Streuung der Ergebnisse auf. Die logistische Regression erreicht als bestes Baseline-Modell einen mittleren ROC-AUC-Wert von 0,798 und liefert damit ein besseres mittleres Ergebnis als alle alleinstehenden ML-Verfahren. Das Modell schlägt in dieser Untersuchung ebenfalls die Ensemble-Verfahren RF und FLAML.

Besser als die logistische Regression schneiden die verbleibenden homogenen und heterogenen Ensemble-Verfahren ab. ASKL erreicht mit 0,810 den höchsten mittleren ROC-AUC-Wert von allen Modellen und weist gleichzeitig die geringste Streuung der Ergebnisse auf.

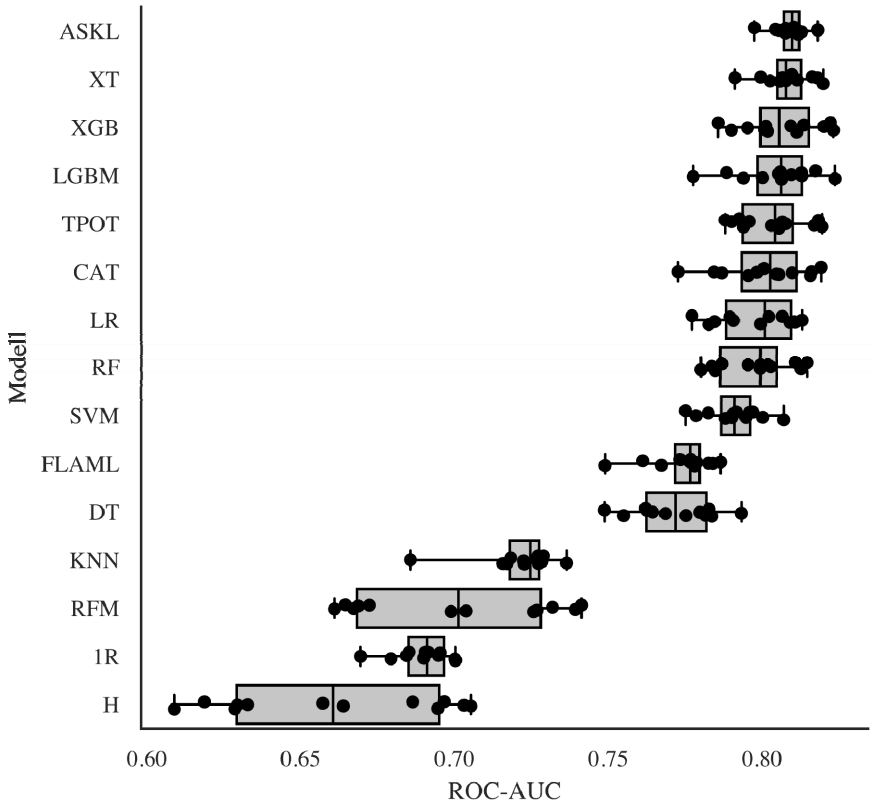

Abbildung 16: Boxplot der ROC-AUC-Ergebnisse über die 12 Testdatenmengen

Die Bewertung anhand des TDL liefert ein in Teilen abweichendes Bild. So erreicht das relative einfache RFM-Modell mit einem mittleren TDL-Wert von 1,745 den vierten Platz im Modellvergleich und muss sich nur den heterogenen Ensembles geschlagen geben, die in Bezug auf diese Metrik die besten mittleren Ergebnisse bei vergleichsweise geringer Streuung liefern. ASKL ist analog zur ROC-AUC-Auswertung das Modell mit dem besten Ergebnis und erreicht einen mittleren TDL-Wert von 1,919.

Die Beurteilung, ob die dargestellten Leistungen der verschiedenen Algorithmen sich statistisch signifikant unterscheiden, basiert auf dem Friedman-Test (vgl. Demšar 2006, S. 11 f. García u. a. 2010, S. 2048 f. Japkowicz, Shah 2011, S. 248 ff.) mit einer Finner-Korrektur für mehrfaches Testen (vgl. Finner 1993; García u. a. 2010, S. 2054 ff.).

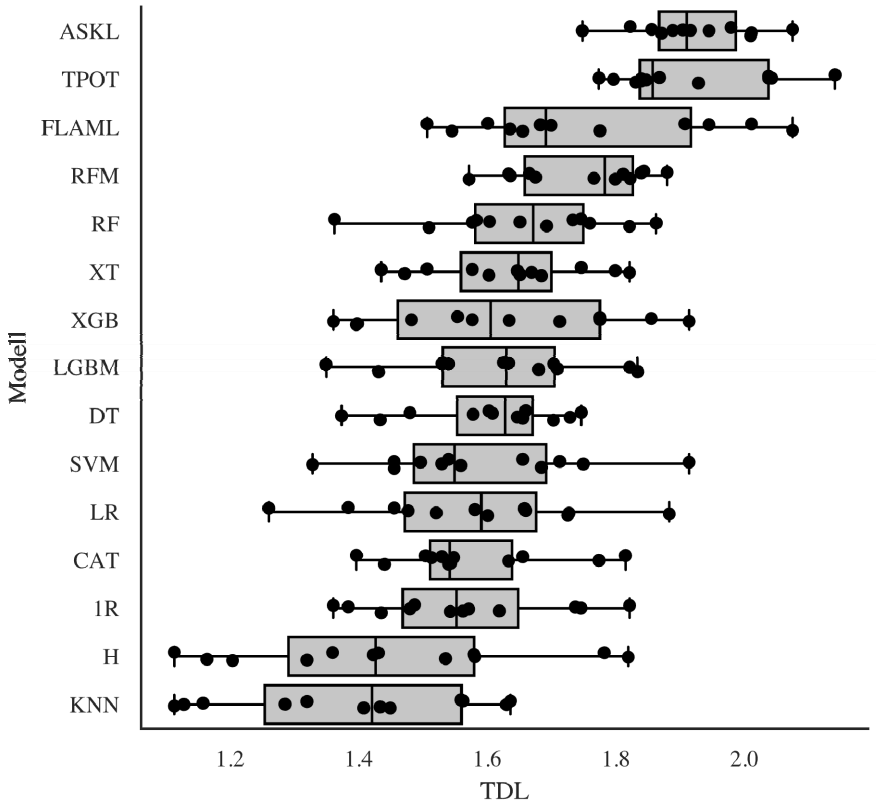

Abbildung 17: Boxplot der TDL-Ergebnisse über die 12 Testdatenmengen

Der Friedman-Test vergleicht die durchschnittlichen Ergebnis-Ränge der Modelle über die verschiedenen Testdatenmengen mit der Nullhypothese, dass diese sich nicht unterscheiden. Im Gesamtvergleich ergeben sich für beide Evaluationsmetriken sehr niedrige p-Werte von $2{,}2*10^{-16}$ (ROC-AUC) sowie $9{,}7*10^{-14}$ (TDL), so dass die Ergebnisunterschiede als statistisch hoch signifikant betrachtet werden können.

Abbildung 18 und Abbildung 19 illustrieren die Ergebnisse der folgenden paarweisen Tests in kompakter Form (vgl. Demšar 2006, S. 15 f.). Die Diagramme ordnen die mittleren Ergebnis-Ränge der verschiedenen Modelle auf einer Achse an, wobei höhere Ränge darauf hinweisen, dass ein Modell durchschnittlich besser abschneidet als Vergleichsmodelle mit niedrigeren Rängen. Zwei oder mehr Modelle sind graphisch mit einem Balken verbunden, falls sich die mittleren Rangunterschiede statistisch nicht zum Niveau 0,05 unterscheiden. Tabelle 5 im Anhang zeigt ergänzend die mittleren Ergebnis-Ränge im Detail.

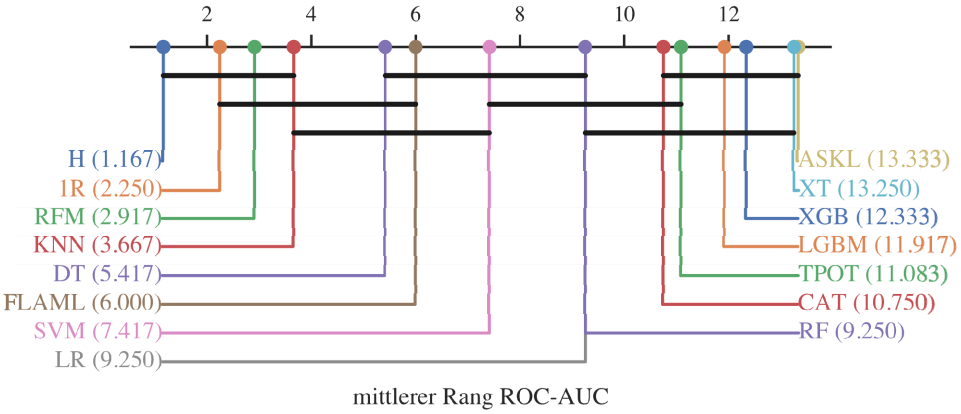

Abbildung 18: Kritische Differenzen ROC-AUC

Das Modell ASKL schneidet sowohl beim ROC-AUC-Wert als auch beim TDL-Wert mit Rängen von 13,333 bzw. 14,333 über alle verschiedenen Testdatenmengen am besten ab.

Bezogen auf den ROC-AUC-Vergleich liegen verschiedene ML-Ensembles auf den folgenden Plätzen, wobei sich die Modell-Leistungen bis einschließlich CAT auf Platz 6 nicht statistisch signifikant unterscheiden. Das bestplatzierte Modell ASKL liefert statistisch hoch signifikant bessere Ergebnisse als alle Baseline-Modelle.

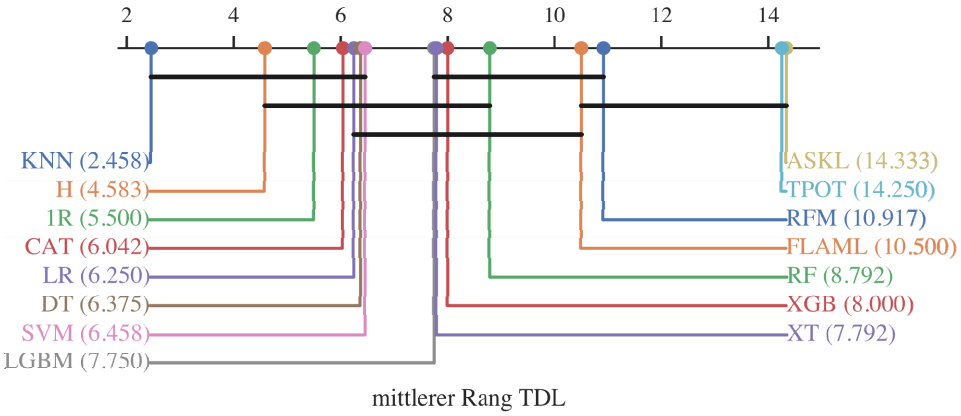

Abbildung 19: Kritische Differenzen TDL

Bei Betrachtung der mittleren Ränge der TDL-Werte landet interessanterweise das einfache RFM-Baseline-Modell auf dem dritten Platz. Die übrigen Plätze unter den ersten vier werden durch die heterogenen Ensembles abgedeckt, wobei die Leistungen der bestplatzierten vier Modelle keine statistisch signifikanten Unterschiede zeigen.

Zusammenfassend liefern ML-Ensembles in Bezug auf beide Evaluationsmetriken die besten Ergebnisse. Mit Ausnahme des RFM-Modells für den TDL sind die Ergebnisunterschiede der besten Ensembles sowohl im Vergleich zu den betrachteten einfachen Baseline-Modellen als auch zu den alleinstehenden ML-Verfahren statistisch hoch signifikant.

Obwohl die Leistungsunterschiede der jeweils besten Modelle nicht statistisch signifikant sind, bietet sich die Auswahl von ASKL als Modell für den produktiven Einsatz an. Das Modell landet bei beiden Bewertungsmetriken auf dem ersten Platz und erlaubt durch die bestehende AutoML-Implementierung eine relativ einfache Konfiguration und Umsetzung in der Praxis.

5.5 Evaluation

5.5.1 Ergebnisse bewerten

Die Evaluation beurteilt, inwieweit das beste Modell die definierten Geschäftsziele erreicht.

Das entwickelte Modell prognostiziert die Abwanderungswahrscheinlichkeit von Kunden signifikant besser als die betrachteten Baselines. Abbildung 20 illustriert die Modell-Leistung anhand einer ROC-Kurve und einer Lift-Kurve jeweils für die aktuellste Testdatenmenge.

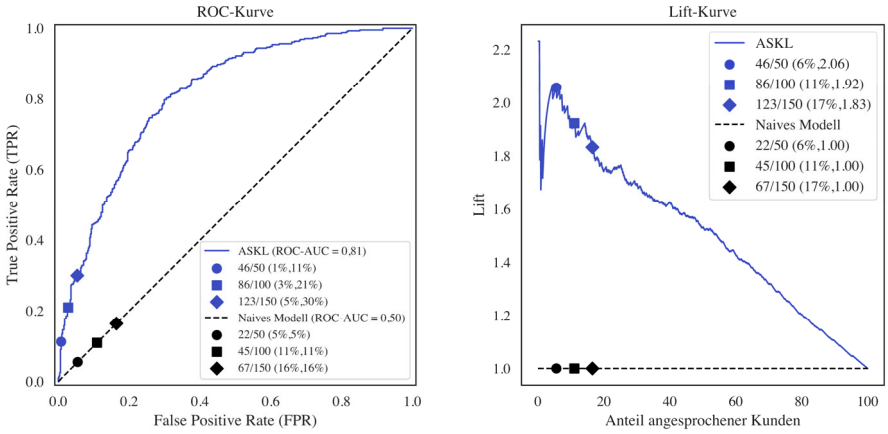

Abbildung 20: Modell-Leistung für ausgewählte Schwellenwerte[1]

[1] Darstellung in Anlehnung an (Gattermann-Itschert, Thonemann 2022, S. 141). Angaben in der Legende: Anzahl korrekter Prognosen / Anzahl ausgewählter Kunden (Wert X-Achse, Wert Y-Achse)

Zum Betrachtungszeitpunkt 31.12.2022 hatte die WER GmbH insgesamt 915 aktive Kunden im Streckengeschäft, wovon 410 oder 44,81% in den folgenden 12 Monaten abgewandert sind.

Das Unternehmen kann nur begrenzt Ressourcen in die Kundenbindung investieren. Die Evaluation betrachtet daher drei realistische Szenarien, in denen jeweils die 50 (●), 100 (■) oder 150 (◆) Kunden mit den höchsten Abwanderungswahrscheinlichkeiten mittels Kundenbindungsmaßnahmen adressiert werden. Abbildung 20 visualisiert die Abwägung zwischen korrekt identifizierten und damit durch Kundenbindungsmaßnahmen potenziell gehaltenen Abwanderern sowie falsch prognostizierten Abwanderern und damit verschwendeten Ressourcen bei den entsprechenden Schwellenwerten. Als Vergleich zum ASKL-Modell ist jeweils der naive Benchmark abgebildet.

Im Fall der Auswahl von 50 Kunden für eine Bindungsmaßnahme identifiziert das Modell 46 reale Abwanderer und erreicht damit eine Präzision der Prognose von 92% bei einer Fehlalarmquote von unter 1%. Das naive Modell würde nur 22 der 50 ausgewählten Kunden korrekt als Abwanderer identifizieren, dies entspricht einer Präzision von 44% bei einer Fehlalarmquote von über 5%. Das entwickelte Modell identifiziert damit bei der Auswahl von 50 Kunden über zweimal mehr Abwanderer (Lift) als der naive Benchmark.

Im Fall der Adressierung von 100 bzw. 150 Kunden sinkt die Präzision des Modells auf 86% bzw. 82% ab, liefert jedoch mit Lift-Werten von 1,92 bzw. 1,93 immer noch signifikant bessere Ergebnisse als der dargestellte naive Vergleichsbenchmark.

5.5.2 Modell interpretieren

Die im vorherigen Abschnitt dargestellte Bewertung der Ergebnisse erlaubt eine Beurteilung der Modell-Leistung. Sie erklärt jedoch nicht, *wie* das Modell die Prognosewerte erzeugt. Die Interpretation des Modells soll dazu beitragen, die Ursachen der Modell-Prognosen zu verstehen bzw. diese zu erklären (vgl. Molnar 2022, S. 13).

Komplexere ML-Verfahren wie das ausgewählte heterogene Ensemble erzeugen i.d.R. „Black-Box"-Modelle, d.h. die Funktionsweise der Ergebnisberechnung ist nicht einfach nachvollziehbar bzw. erklärbar.

Ein verbreitetes und leistungsfähiges, modell-agnostisches Verfahren zur Interpretation verschiedenster Modelltypen besteht in der Berechnung und Analyse von Shapley-Werten (engl.: Shapley Values) (vgl. Lundberg, Lee 2017; Molnar 2022, S. 215 ff.). Shapley-Werte werden in einem spieltheoretischen Ansatz ermittelt und quantifizieren den durchschnittlichen Beitrag einzelner Merkmalswerte zur Modellausgabe. Sie liefern Erkenntnisse sowohl auf lokaler Ebene (pro Datensatz) als auch auf globaler Ebene (datensatzübergreifend).

Die globale Bedeutung einzelner Merkmale kann über die mittleren absoluten Shapley-Werte je Merkmal interpretiert werden. Abbildung 21 visualisiert die so berechnete

Merkmalsbedeutung für die 10 Merkmale mit der durchschnittlich höchsten absoluten Auswirkung auf die Prognose des ASKL-Modells[2].

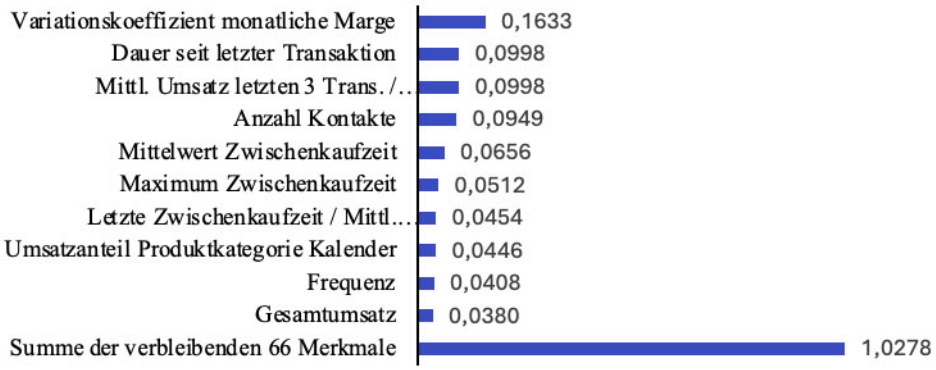

Abbildung 21: Globale Merkmalsbedeutung

Abbildung 22 ergänzt die globale Merkmalsbedeutung um Informationen zur Wirkungsrichtung. Die Visualisierung zeigt die berechneten Shapley-Werte für alle Instanzen des aktuellsten Testdatensatzes, wobei die Position auf der y-Achse dem betrachteten Merkmal und die Position auf der x-Achse dem Shapley-Wert auf einer Log-Odds-Skala entspricht. Die Farbe repräsentiert den Wert des jeweiligen Merkmals von gering bis hoch, womit die Darstellung eine grobe Interpretation des Zusammenhangs zwischen Merkmalswert und Ergebnisauswirkung unterstützt.

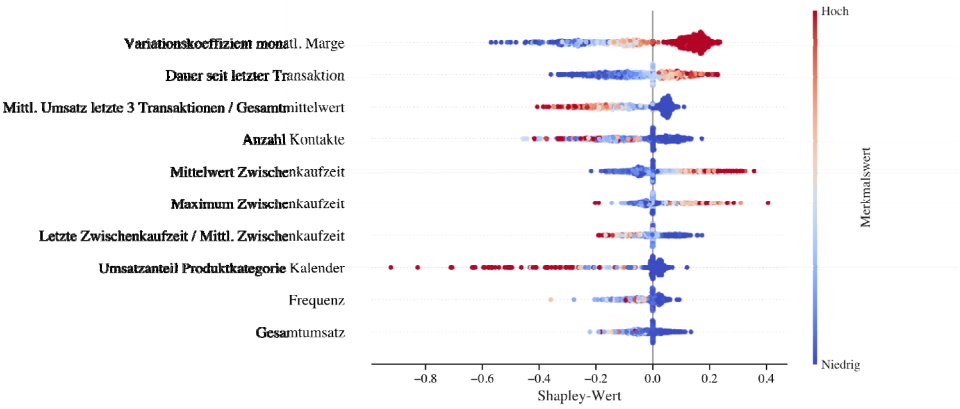

Abbildung 22: Shapley-Werte der wichtigsten Merkmale

[2] Basis der Darstellung sind alle Instanzen des aktuellsten Testdatensatzes. Die Shapley-Werte sind auf einer Log-Odds-Skala dargestellt.

Die Visualisierungen liefern eine Reihe von Einblicken in die Funktionsweise des Prognosemodells. Die transaktionalen RFM-Merkmale stellen mit einem Anteil von 8 den Großteil der für das Modell bedeutendsten 10 Merkmale, was die Bedeutung dieser Merkmalskategorie für die Abwanderungsprognose unterstreicht.

Das Merkmal mit der höchsten globalen Bedeutung ist der Variationskoeffizient der monatlichen Marge. Das Modell interpretiert hierbei Kunden mit einem unregelmäßigeren monatlichen Ergebnisbeitrag als tendenziell abwanderungsgefährdeter. Dieser beobachtete Einfluss der Schwankung monatlicher Ergebnisbeiträge deckt sich mit Ergebnissen aus vergleichbaren Untersuchungen im B2B-Bereich (vgl. Gattermann-Itschert, Thonemann 2022, S. 140 f.; Mirkovic u. a. 2022, S. 12 f.).

Weitere bedeutende monetäre Merkmale betreffen die Entwicklungstendenz sowie die absolute Höhe der Umsätze: Kunden, deren mittlerer Auftragswert bezogen auf die letzten drei Aufträge über dem Gesamtmittelwert der Aufträge liegt, werden vom Modell als weniger abwanderungsgefährdet eingestuft. Eine ähnliche Logik greift beim Gesamtumsatz der Kundenbeziehung: je höher dieser desto tendenziell geringer die Abwanderungswahrscheinlichkeit.

Das zweitwichtigste Merkmal ist die Recency, d.h. die Dauer seit der letzten Transaktion in Tagen. Abbildung 23 stellt den Zusammenhang zwischen Recency- und Shapley-Werten detaillierter dar: das Modell betrachtet einen Kunden als umso stärker gebunden, je weniger Tage sein letzter Auftrag zurück liegt. Eine Art von „Wendepunkt“ liegt bei ca. 150 Tagen: geringere Recency-Werte reduzieren tendenziell die Abwanderungswahrscheinlichkeit während höhere Recency-Werte diese tendenziell erhöhen.

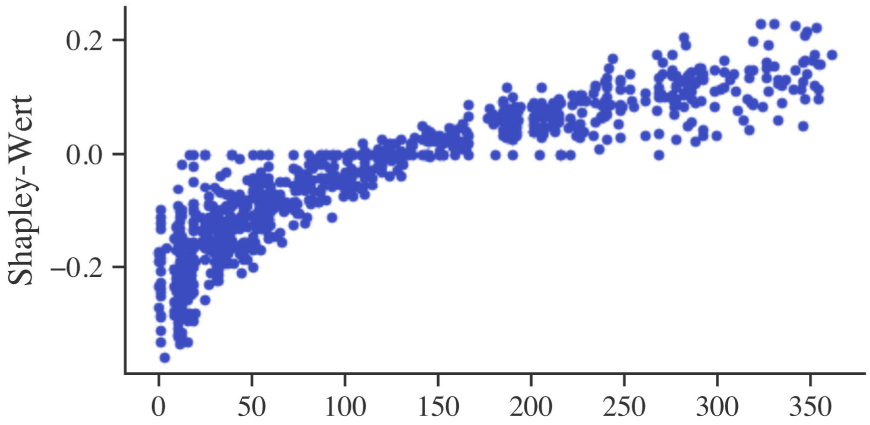

Abbildung 23: Merkmalsabhängigkeit Recency

Im Bereich der Transaktionshäufigkeiten sind für das Modell sowohl die Gesamtanzahl der Transaktionen als auch verschiedene Aggregationen der Zwischenkaufzeiten bedeutsam. Je häufiger ein Kunde bereits beim Unternehmen eingekauft hat, desto geringer wird die Abwanderungswahrscheinlichkeit eingestuft. Je länger die Zeiten zwischen zwei Aufträgen – sowohl im Mittel als auch im Maximum – sind, desto abwanderungsgefährdeter sieht das Modell den jeweiligen Kunden. Eine Verkürzung der Zwischenkaufzeit im Zeitablauf - betrachtet als Verhältnis von letzter Zwischenkaufzeit zur mittleren Zwischenkaufzeit - führt im Modell zu einer sinkenden Abwanderungswahrscheinlichkeit.

Aus dem Bereich der interaktionsbezogenen Merkmale fällt die Gesamtzahl der Kundenkontakte als bedeutendes Merkmal auf: je häufiger ein Kunde in der Vergangenheit durch den Vertrieb des Unternehmens kontaktiert wurde, desto geringer die prognostizierte Abwanderungswahrscheinlichkeit.

Eine interessant hohe Bedeutung hat der relative Umsatzanteil an der Produktkategorie „Kalender“ aus dem Bereich der leistungs- bzw. angebotsbezogenen Merkmale: je höher dieser Anteil, desto stärker senkt das Modell die prognostizierte Abwanderungswahrscheinlichkeit ab (vgl. Abbildung 24). Dieser Zusammenhang könnte darauf hindeuten, dass Kunden Kalender jährlich wiederholend kaufen oder ein Indiz für den positiven Zusammenhang zwischen Cross-Selling bzw. Angebotsbreite und Kundenbindung sein.

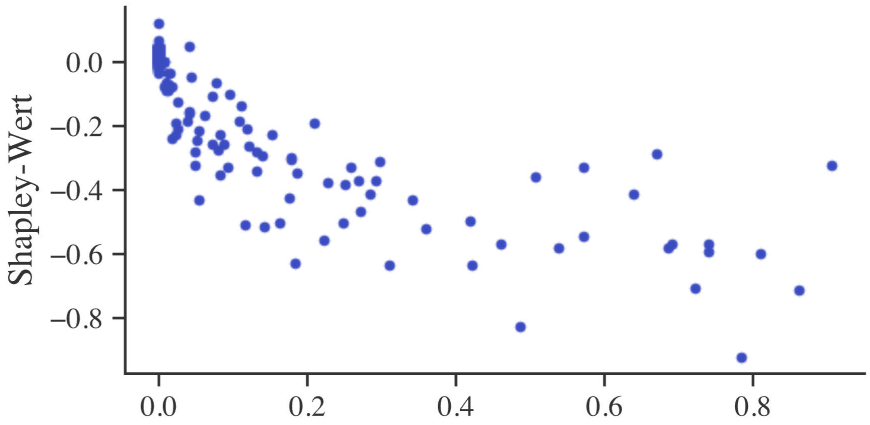

Abbildung 24: Merkmalsabhängigkeit Umsatzanteil Produktkategorie Kalender

5.6 Bereitstellung

Die Bereitstellung integriert das entwickelte Modell in die Organisation und IT-Landschaft der WER GmbH.

Die in Abschnitt 5 beschriebenen Funktionen zur Datensammlung, Datenaufbereitung und Modellbildung sind in Programmen implementiert und werden periodisch nach dem Abschluss eines Buchungsmonats automatisiert ausgeführt. Basierend auf dem jeweils auf den aktuellsten Daten trainierten Modell werden die Score-Werte bzw. Abwanderungswahrscheinlichkeiten für alle Kunden berechnet und in einer Controlling-Plattform auf Basis von Microsoft Power BI hinterlegt.

Zur Planung, Steuerung und Kontrolle von Kundenbindungsmaßnahmen entwickelte das Unternehmen ein Power BI-Dashboard zur Nutzung durch Vertriebsmitarbeiter und die Geschäftsführung (vgl. Lamsayah 2023). Das Dashboard setzt Kunden automatisiert aufgrund verschiedener, regelbasierter Kriterien auf eine „Watchlist“ für die Bearbeitung mittels verschiedener Kundenbindungsmaßnahmen. Als ein Kriterium für die Aufnahme eines Kunden auf die Watchlist dient hierbei eine prognostizierte Abwanderungswahrscheinlichkeit von über 50% bei einem gleichzeitig separat abgeschätzten Umsatzpotenzial von EUR 5‘000. Abbildung 25 illustriert den Aufbau des Dashboards überblicksartig.

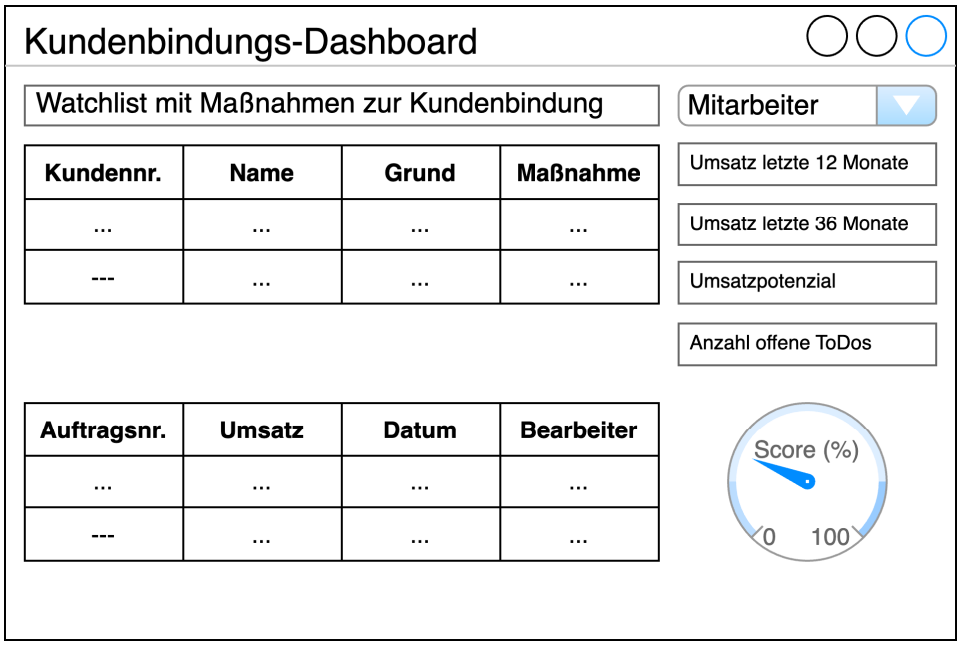

Abbildung 25: Skizze Kundenbindungs-Dashboard

Das Dashboard erlaubt die Auswahl eines oder mehrere Vertriebsmitarbeiter, woraufhin die Watchlist auf die dem ausgewählten Mitarbeiter zugeordneten Kunden eingeschränkt wird. Die Watchlist selbst enthält neben Kundennummer und -name eine Auflistung der Kriterien, warum der Kunden aufgeführt ist und einen ebenfalls regelbasiert erstellten Vorschlag für eine durchzuführende Kundenbindungsmaßnahme.

Mit Auswahl eines Kunden zeigt das Dashboard zum einen im unteren Teil Detailinformationen zu den letzten Transaktionen des Kunden auf Auftragsebene. Darüber hinaus stellt das Dashboard im rechten Teil ausgewählte Kennzahlen zum Kunden dar. Neben dem Umsatz der letzten 12 und 36 Monate sowie einer Abschätzung des Umsatzpotenzials wird hier prominent die durch das entwickelte Modell prognostizierte Abwanderungswahrscheinlichkeit dargestellt. Über eine Drill-Through-Funktion kann der Benutzer in weitere, detailliertere Berichtsansichten zum Kunden abspringen.

Eine Eingabefunktionalität auf einzelnen Zeilen der Watchlist erlaubt die Rückmeldung über die Durchführung der vorgeschlagenen Bindungsmaßnahmen und ermöglicht somit eine Abarbeitungssteuerung für den Benutzer und liefert eine Controlling-Möglichkeit für die Geschäftsführung.

6 Zusammenfassung und Ausblick

6.1 Zusammenfassung

Die WER GmbH verfolgte das Ziel, die Kundenbindung im Streckengeschäft durch ein proaktives Kundenbindungsmanagement systematisch zu reduzieren. Zu diesem Zweck entwickelte die Arbeit ein datengestütztes Modell, dass die Abwanderungswahrscheinlichkeit von Kunden präzise vorhersagt und damit als ein elementarer Baustein für die passgenaue Adressierung von Kunden mit Bindungsmaßnahmen dient. Der Schwerpunkt der Arbeit lag hierbei auf der Entwicklung von kombinierten Modellen – sog. Ensembles – aus dem Bereich des maschinellen Lernens (ML).

Die Arbeit verglich in einem systematischen Test-Design und unter Betrachtung zweier ausgewählter Bewertungsmetriken insgesamt 10 aktuelle ML-Verfahren sowohl untereinander als auch in Relation zu 5 einfacheren Baseline-Verfahren. Das final ausgewählte Modell lieferte signifikant bessere Prognosen als sämtliche Baseline-Verfahren und zeigte damit einen klaren Mehrwert der effektiveren Datennutzung auf. Die Interpretation des Modells über Shapley-Werte lieferte Einblicke in die Bedeutung wesentlicher Merkmale und lieferte Beiträge zur Plausibilisierung der Funktionsweise des Modells.

Die Integration der Modellergebnisse in ein Controlling-Dashboard für das Kundenbindungsmanagement illustrierte die konkrete Nutzung des Modells in der betrieblichen Praxis.

6.2 Ausblick

Eine Überprüfung bestehender Schwächen der Arbeit liefert eine Reihe von Ansatzpunkten für mögliche Folgearbeiten in drei identifizierten Bereichen.

Ein erster Bereich betrifft die Bewertung der Modell-Leistung, welche ausschließlich auf den prognostizierten Abwanderungswahrscheinlichkeiten basiert. Diese sind für sich allein genommen noch kein ausreichendes Kriterium für die kampagnenorientierte Auswahl von Kunden, weshalb das entwickelte Dashboard weitere Informationen als Entscheidungsgrundlage beisteuert. Die Entscheidungsgrundlage könnte zukünftig verbessert werden, indem das Modell um kosten-nutzen-basierte Informationen ergänzt wird. Sind bspw. geplante Kundenbindungsmaßnahmen a priori bekannt, könnten entsprechende Kosten- und Nutzenerwartungen in die Leistungsbewertung der Modelle oder auch in die Modellbildung selbst integriert werden und damit sowohl die Auswahl der Kunden vereinfachen als auch den erwarteten Ergebnisbeitrag erhöhen (vgl. Janssens u. a. 2022; Lemmens, Gupta 2020; Verbraken u. a. 2013).

In einem zweiten Bereich könnte die Datenbasis als Grundlage der Prognose um zusätzliche Merkmale ergänzt werden. Die verwendetet Datengrundlage basiert auf vglw. einfach verfügbaren Daten aus den unternehmensinternen Transaktionssystemen. Eine Ausweitung möglicher Prädiktoren bspw. durch Einbezug von Daten aus Newsletter-Systemen, die systematische Erfassung von Reaktionen auf Marktbearbeitungsmaßnahmen oder die Beschaffung von Wettbewerbssituationen könnte die Prognoseleistung weiter steigern.

Ein dritter Bereich schließlich betrifft die Evaluation von Deep-Learning-Verfahren bzw. tiefen neuronalen Netzen. Diese wurden in der Untersuchung zunächst ausgeschlossen, da auf den konstruierten tabellarischen Trainingsdaten i.d.R. keine bessere Leistungs-/Aufwands-Relation im Vergleich zu ML-Ensembles zu erwarten ist. Ein interessanter Untersuchungsbereich besteht allerdings ggf. im Einsatz von rekurrenten Netzwerkarchitekturen (vgl. Mena u. a. 2024). Diese erlauben die direkte Verarbeitung von im Zeitablauf variierenden Merkmalen (vgl. Abschnitt 5.3.2), ohne dass diese vorab – unter Informationsverlust – aggregiert werden müssen. Hierdurch könnte zum einen der Aufwand der Datenaufbereitung reduziert werden. Zum anderen könnten die Modelle ggf. weitergehende Muster aus den Daten extrahieren und dadurch eine höhere Modell-Leistung erzielen.

Anhang

Tabelle 3: Konstruierte Merkmale

Kategorie	**Unter-kategorie**	**Merkmal**	**Aggregationen**	**Anzahl**
Transaktionsbezogen	Aktualität	Dauer seit der letzten Transaktion in Tagen	-	1
	Frequenz	Anzahl Transaktionen im Betrachtungszeit-raum	-	1
		Anzahl Transaktionen im letzten Monat	-	1
		Anzahl Transaktionen im letzten Quartal	-	1
		Anzahl Transaktionen im letzten Jahr	-	1
		Anzahl Transaktionen / Beziehungsdauer	-	1
		Zwischenkaufzeit (Tage)	Minimum, Maximum, Mittelwert, Median, Standardabweichung, Variationskoeffizient, Slope, letzter Wert / Mittelwert, Mittel letzter 3 Werte / Mittelwert	9
		Recency / mittlere Zwischenkaufzeit	-	1
		Recency / Median Zwischenkaufzeit	-	1
	Monetär	Umsatz je Transaktion	Minimum, Maximum, Mittelwert, Median, Standardabweichung,	9

			Variationskoefffizient, Slope, Letzter Wert / Mittelwert, Mittel letzter 3 Werte / Mittelwert	
		Umsatz je Monat	Minimum, Maximum, Mittelwert, Median, Standardabweichung, Variationskoefffizient, Slope, Letzter Wert / Mittelwert, Mittel letzter 3 Werte / Mittelwert	9
		Gesamtumsatz	-	1
		Gesamtumsatz / Beziehungsdauer	-	1
		Umsatz letzter Monat	-	1
		Umsatz letztes Quartal	-	1
		Umsatz letztes Jahr	-	1
		Marge je Transaktion	Minimum, Maximum, Mittelwert, Median, Standardabweichung, Variationskoefffizient, Slope, Letzter Wert / Mittelwert, Mittel letzter 3 Werte / Mittelwert	9
		Marge je Monat	Minimum, Maximum, Mittelwert, Median, Standardabweichung, Variationskoefffizient, Slope, Letzter Wert / Mittelwert, Mittel letzter 3 Werte / Mittelwert	9
		Gesamtmarge	-	1
		Gesamtmarge / Beziehungsdauer	-	1

		Marge letzter Monat	-	1
		Marge letztes Quartal	-	1
		Marge letztes Jahr	-	1
Leistungs- bzw. Angebotsbezogen	-	Relativer Umsatzanteil je Produktkategorie	-	5
		Anzahl jemals gekaufter Produktkategorien	-	1
Kundenbezogen	-	Ort des Kunden (PLZ-Bereich)	-	1
	-	Organisch vs. zugekauft		1
Interaktionsbezogen	-	Beziehungsdauer	-	1
		Anzahl Kontakte	-	1
		Tage seit letztem Kontakt	-	1
		Art des letzten Kontakts	-	1
		Zugeordneter Vertriebsmitarbeiter	-	1
				76

Tabelle 4: Evaluationsergebnisse (Metriken)

Klasse	Modell	AUC		TDL	
		Mittelwert	**Standard-abweichung**	**Mittelwert**	**Standard-abweichung**
Baseline	H	0,660	0,036	1,443	0,227
	RFM	0,700	0,032	1,745	0,103
	1R	0,690	0,009	1,562	0,147
	LR	0,798	0,013	1,578	0,170
Alleinstehend	DT	0,771	0,013	1,601	0,117
	KNN	0,721	0,013	1,390	0,190
	SVM	0,791	0,009	1,590	0,159
Homogenes Ensemble	RF	0,798	0,012	1,659	0,141
	XT	0,808	0,008	1,634	0,123
	XGB	0,806	0,012	1,620	0,188
	LGBM	0,805	0,013	1,615	0,147
	CAT	0,801	0,014	1,574	0,125
Heterogenes Ensemble	ASKL	0,810	0,006	1,919	0,091
	FLAML	0,774	0,011	1,753	0,188
	TPOT	0,803	0,011	1,915	0,119

Tabelle 5: Evaluationsergebnisse (Ränge)

Klasse	**Modell**	**mittlerer Rang ROC-AUC**	**mittlerer Rang TDL**
Baseline	H	1,167	4,583
	RFM	2,917	10,917
	1R	2,250	5,500
	LR	9,250	6,250
Alleinstehend	DT	5,417	6,375
	KNN	3,667	2,458
	SVM	7,417	6,458
Homogenes Ensemble	RF	9,250	8,792
	XT	13,250	7,792
	XGB	12,333	8,000
	LGBM	11,917	7,750
	CAT	10,750	6,042
Heterogenes Ensemble	ASKL	13,333	14,333
	FLAML	6,000	10,500
	TPOT	10,083	14,250

Literaturverzeichnis

Ascarza, Eva; Neslin, Scott A.; Netzer, Oded; Anderson, Zachery; Fader, Peter S.; Gupta, Sunil; Hardie, Bruce G. S.; Lemmens, Aurélie; Libai, Barak; Neal, David; Provost, Foster und Schrift, Rom, 2018. In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions. *Customer Needs and Solutions*. März 2018. Bd. 5, Nr. 1–2, S. 65–81. DOI 10.1007/s40547-017-0080-0

Bergstra, James; Bardenet, Rémi; Bengio, Yoshua und Kégl, Balázs, 2011. Algorithms for hyper-parameter optimization. In: *Proceedings of the 24th international conference on neural information processing systems*. Red Hook, NY, USA: Curran Associates Inc. 2011. S. 2546–2554. NIPS'11. ISBN 978-1-61839-599-3

Berthold, Michael R.; Borgelt, Christian; Höppner, Frank; Klawonn, Frank und Silipo, Rosaria, 2020. *Guide to Intelligent Data Science: How to Intelligently Make Use of Real Data*. Cham: Springer International Publishing. Texts in Computer Science. ISBN 978-3-030-45573-6

Boenigk, Silke, 2011. Kündigungspräventionsmanagement. In: Hippner, Hajo; Hubrich, Beate und Wilde, Klaus D. (Hrsg.), *Grundlagen des CRM*. Wiesbaden: Gabler. S. 475–497. ISBN 978-3-8349-2550-3

Bogaert, Matthias und Delaere, Lex, 2023. Ensemble Methods in Customer Churn Prediction: A Comparative Analysis of the State-of-the-Art. *Mathematics*. 24 Februar 2023. Bd. 11, Nr. 5, S. 1137. DOI 10.3390/math11051137

Borisov, Vadim; Leemann, Tobias; Seßler, Kathrin; Haug, Johannes; Pawelczyk, Martin und Kasneci, Gjergji, 2024. Deep Neural Networks and Tabular Data: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*. 2024. Bd. 35, Nr. 6, S. 7499–7519. DOI 10.1109/TNNLS.2022.3229161

Breiman, Leo, 2001a. Statistical Modeling: The Two Cultures. *Statistical Science*. 2001. Bd. 16, Nr. 3, S. 199–215

Breiman, Leo, 2001b. Random Forests. *Machine Learning*. 2001. Bd. 45, Nr. 1, S. 5–32. DOI 10.1023/A:1010933404324

Brownlee, Jason, 2019. Information Gain and Mutual Information for Machine Learning. *MachineLearningMastery.com* [online]. 15 Oktober 2019. [Zugriff am: 26 Mai 2024]. Verfügbar unter: https://machinelearningmastery.com/information-gain-and-mutual-information/

Bruhn, Manfred, 2022. *Relationship Marketing: das Management von Kundenbeziehungen*. 6., überarbeitete Auflage. München: Verlag Franz Vahlen. Vahlens Handbücher der Wirtschafts- und Sozialwissenschaften. ISBN 978-3-8006-6622-5

Buckinx, Wouter und Van den Poel, Dirk, 2005. Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*. Juli 2005. Bd. 164, Nr. 1, S. 252–268. DOI 10.1016/j.ejor.2003.12.010

Chapman, Pete; Clinton, Julian; Kerber, Randy; Khabaza, Thomas; Reinartz, Thomas; Shearer, Colin und Wirth, Rüdiger, 2010. *CRISP-DM 1.0: Step-by-step data mining guide*. CRISP-DM Consortium

Chen, Tianqi und Guestrin, Carlos, 2016. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco California USA: ACM. 13 August 2016. S. 785–794. ISBN 978-1-4503-4232-2

Demšar, Janez, 2006. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*. 2006. Bd. 7, Nr. 1, S. 1–30

Diller, Hermann, 1995. *Kundenbindung als Zielvorgabe im Beziehungs-Marketing*. Arbeitspapier Nr. 40. Nürnberg: Universität Erlangen-Nürnberg, Betriebswirtschaftliches Institut, Lehrstuhl für Marketing. Arbeitspapiere des Lehrstuhls für Marketing an der Universität Erlangen-Nürnberg

Fader, Peter S. und Hardie, Bruce G.S., 2009. Probability Models for Customer-Base Analysis. *Journal of Interactive Marketing*. Februar 2009. Bd. 23, Nr. 1, S. 61–69. DOI 10.1016/j.intmar.2008.11.003

Feurer, Matthias und Hutter, Frank, 2019. Hyperparameter Optimization. In: Hutter, Frank; Kotthoff, Lars und Vanschoren, Joaquin (Hrsg.), *Automated Machine Learning*. Cham: Springer International Publishing. S. 3–33. The Springer Series on Challenges in Machine Learning. ISBN 978-3-030-05317-8

Feurer, Matthias; Klein, Aaron; Eggensperger, Katharina; Springenberg, Jost; Blum, Manuel und Hutter, Frank, 2015. Efficient and robust automated machine learning. In: *Advances in neural information processing systems*. 2015. S. 2962–2970

Finner, H., 1993. On a Monotonicity Problem in Step-Down Multiple Test Procedures. *Journal of the American Statistical Association*. September 1993. Bd. 88, Nr. 423, S. 920–923. DOI 10.1080/01621459.1993.10476358

Flach, Peter A., 2011. ROC Analysis. In: Sammut, Claude und Webb, Geoffrey I. (Hrsg.), *Encyclopedia of Machine Learning*. Boston, MA: Springer US. S. 869–875. ISBN 978-0-387-30768-8

Friedman, Jerome H., 2001. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*. 1 Oktober 2001. Bd. 29, Nr. 5. DOI 10.1214/aos/1013203451

Friedman, Jerome; Hastie, Trevor und Tibshirani, Robert, 2009. *The Elements of Statistical Learning*. Second edition, corrected at 12th printing 2017. New York, NY: Springer. Springer series in statistics. ISBN 978-0-387-84858-7

García, Salvador; Fernández, Alberto; Luengo, Julián und Herrera, Francisco, 2010. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Information Sciences*. Mai 2010. Bd. 180, Nr. 10, S. 2044–2064. DOI 10.1016/j.ins.2009.12.010

Gattermann-Itschert, Theresa und Thonemann, Ulrich W., 2021. How training on multiple time slices improves performance in churn prediction. *European Journal of Operational Research*. Dezember 2021. Bd. 295, Nr. 2, S. 664–674. DOI 10.1016/j.ejor.2021.05.035

Gattermann-Itschert, Theresa und Thonemann, Ulrich W., 2022. Proactive customer retention management in a non-contractual B2B setting based on churn prediction with random forests. *Industrial Marketing Management*. November 2022. Bd. 107, S. 134–147. DOI 10.1016/j.indmarman.2022.09.023

Geiger, 2024. *Homepage Geiger* [online]. 2024. [Zugriff am: 23 August 2024]. Verfügbar unter: https://geiger.com

Géron, Aurélien, 2022. *Hands-On machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems*. Third edition. Beijing; Boston; Tokyo: O'Reilly. ISBN 978-1-09-812596-7

Geurts, Pierre; Ernst, Damien und Wehenkel, Louis, 2006. Extremely randomized trees. *Machine Learning*. April 2006. Bd. 63, Nr. 1, S. 3–42. DOI 10.1007/s10994-006-6226-1

Gijsbers, Pieter; Bueno, Marcos L. P.; Coors, Stefan; LeDell, Erin; Poirier, Sébastien; Thomas, Janek; Bischl, Bernd und Vanschoren, Joaquin, 2024. AMLB: an AutoML Benchmark. *Journal of Machine Learning Research*. 2024. Bd. 25, Nr. 101, S. 1–65

Grinsztajn, Leo; Oyallon, Edouard und Varoquaux, Gael, 2022. Why do tree-based models still outperform deep learning on typical tabular data? In: Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K. und Oh, A. (Hrsg.), *Advances in Neural Information Processing Systems*. Curran Associates, Inc. 2022. S. 507–520

Gupta, Sunil; Hanssens, Dominique; Hardie, Bruce; Kahn, Wiliam; Kumar, V.; Lin, Nathaniel; Ravishanker, Nalini und Sriram, S., 2006. Modeling Customer Lifetime Value. *Journal of Service Research*. November 2006. Bd. 9, Nr. 2, S. 139–155. DOI 10.1177/1094670506293810

Gupta, Sunil; Lehmann, Donald R. und Stuart, Jennifer Ames, 2004. Valuing Customers. *Journal of Marketing Research*. 2004. Bd. 41, Nr. 1, S. 7–18 JSTOR

Gür Ali, Özden und Arıtürk, Umut, 2014. Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. *Expert Systems with Applications*. Dezember 2014. Bd. 41, Nr. 17, S. 7889–7903. DOI 10.1016/j.eswa.2014.06.018

GWW, 2023. *Werbeartikel-Monitor 2023*. 2023. Gesamtverband der Werbeartikel-Wirtschaft e.V.

GWW, 2024. *Werbeartikel-Monitor 2024*. 2024. Gesamtverband der Werbeartikel-Wirtschaft e.V.

Heinemann, Gerrit, 2022. *Der neue Online-Handel: Geschäftsmodelle, Geschäftssysteme und Benchmarks im E-Commerce*. Wiesbaden: Springer Fachmedien Wiesbaden. ISBN 978-3-658-36664-3

Helander, Mary E., 2021. Solution Methodologies. In: Cochran, James J. (Hrsg.), *INFORMS Analytics Body of Knowledge*. INFORMS. S. 99–154. ISBN 978-0-9906153-6-1

Holte, Robert C., 1993. Very Simple Classification Rules Perform Well on Most Commonly Used Datasets. *Machine Learning*. 1993. Bd. 11, Nr. 1, S. 63–90. DOI 10.1023/A:1022631118932

Hyndman, Rob J. und Athanasopoulos, George, 2021. Time series cross-validation. In: *Forecasting: Principles and Practice* [online]. 3. Melbourne, Australia: OTexts. [Zugriff am: 3 März 2021]. Verfügbar unter: https://otexts.com/fpp3/

Iba, Wayne und Langley, Pat, 1992. Induction of One-Level Decision Trees. In: Sleeman, Derek und Ewards, Peter (Hrsg.), *Machine Learning Proceedings*. San Mateo: Morgan Kaufmann. S. 233–240. ISBN 978-1-55860-247-2

Janssens, Bram; Bogaert, Matthias; Bagué, Astrid und Van Den Poel, Dirk, 2022. B2Boost: instance-dependent profit-driven modelling of B2B churn. *Annals of Operations Research*. 16 März 2022. Bd. 341. DOI 10.1007/s10479-022-04631-5

Januschowski, Tim; Gasthaus, Jan; Wang, Yuyang; Salinas, David; Flunkert, Valentin; Bohlke-Schneider, Michael und Callot, Laurent, 2020. Criteria for classifying forecasting methods. *International Journal of Forecasting*. Januar 2020. Bd. 36, Nr. 1, S. 167–177. DOI 10.1016/j.ijforecast.2019.05.008

Japkowicz, Nathalie und Shah, Mohak, 2011. *Evaluating Learning Algorithms: a classification perspective*. Cambridge ; New York: Cambridge University Press. ISBN 978-0-521-19600-0

Kalisch, Markus und Meier, Lukas, 2021. *Logistische Regression: Eine anwendungsorientierte Einführung mit R*. Wiesbaden: Springer Fachmedien Wiesbaden. essentials. ISBN 978-3-658-34224-1

Karmaker, Shubhra Kanti; Hassan, Md. Mahadi; Smith, Micah J.; Xu, Lei; Zhai, Chengxiang und Veeramachaneni, Kalyan, 2022. AutoML to Date and Beyond: Challenges and Opportunities. *ACM Computing Surveys*. 30 November 2022. Bd. 54, Nr. 8, S. 1–36. DOI 10.1145/3470918

Ke, Guolin; Meng, Qi; Finley, Thomas; Wang, Taifeng; Chen, Wei; Ma, Weidong; Ye, Qiwei und Liu, Tie-Yan, 2017. LightGBM: A highly efficient gradient boosting decision tree. In: Von Luxburg, Ulrike; Guyon, Isabelle; Bengio, Samy; Wallach, Hanna und Fergus, Rob (Hrsg.), *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc. 2017. S. 3146–3154. ISBN 978-1-5108-6096-4

Kotler, Philip; Armstrong, Gary; Harris, Lloyd C und Piercy, Nigel, 2019. *Grundlagen des Marketing*. Hallbergmoos: Pearson. ISBN 978-3-86326-850-3

Kuhn, Max und Johnson, Kjell, 2020. *Feature engineering and selection: a practical approach for predictive models*. Boca Raton London New York: CRC Press, Taylor & Francis Group. Chapman & Hall/CRC data science series. ISBN 978-1-138-07922-9

Kumar, V. und Reinartz, Werner, 2018. *Customer Relationship Management*. Berlin, Heidelberg: Springer Berlin Heidelberg. Springer Texts in Business and Economics. ISBN 978-3-662-55380-0

Lamsayah, Mohamed, 2023. *Konzeption von Kundenbindungsmaßnahmen und Implementierung eines Controlling-Cockpits im Werbemittelhandel am Beispiel der WER GmbH*. Master-Thesis. Düsseldorf: Hochschule Düsseldorf

Lemmens, Aurelie und Gupta, Sunil, 2020. Managing Churn to Maximize Profits. *Marketing Science*. 2020. Bd. 39, Nr. 5, S. 956–973. DOI 10.2139/ssrn.2964906

Lundberg, Scott M. und Lee, Su-In, 2017. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc. 2017. S. 4768–4777. NIPS'17. ISBN 978-1-5108-6096-4

Mariscal, Gonzalo; Marbán, Óscar und Fernández, Covadonga, 2010. A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*. Juni 2010. Bd. 25, Nr. 2, S. 137–166. DOI 10.1017/S0269888910000032

Mena, Gary; Coussement, Kristof; De Bock, Koen W.; De Caigny, Arno und Lessmann, Stefan, 2024. Exploiting time-varying RFM measures for customer churn prediction with deep neural networks. *Annals of Operations Research*. 2024. Bd. 339, Nr. 1, S. 765–787. DOI 10.1007/s10479-023-05259-9

Mirkovic, Milan; Lolic, Teodora; Stefanovic, Darko; Anderla, Andras und Gracanin, Danijela, 2022. Customer churn prediction in B2B non-contractual business settings using invoice data. *Applied Sciences*. 2022. Bd. 12, Nr. 10. DOI 10.3390/app12105001

Molnar, Christoph, 2022. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Second Edition. Victoria, British Columbia, Canada: Leanpub. ISBN 9798411463330

Murphy, Kevin P., 2022. *Probabilistic Machine Learning: An Introduction*. Cambridge, Massachusetts London, England: The MIT Press. Adaptive computation and machine learning. ISBN 978-0-262-04682-4

Naß, Stephan, 2012. *Strategisches Kündigungsverhalten: Eine empirische Analyse vertraglicher Endkundenbeziehungen*. Wiesbaden: Gabler Verlag. Kundenmanagement & Electronic Commerce. ISBN 978-3-8349-6755-8
658

Nebert, Michael, 2006. *Praxishandbuch Werbeartikel: clever briefen, Einsparpotenziale nutzen, Vertriebswege auswählen, effektiv streuen, Erfolgskontrolle sichern*. 2., aktualisierte und erw. Aufl. Heidelberg: MI-Fachverl., Redline. ISBN 978-3-636-03072-6

Nielsen, 2024. *Entwicklung der Bruttoinvestitionen in Werbung in Deutschland in den Jahren 1995 bis 2023* [online]. 18 Januar 2024. Statista. [Zugriff am: 9 August 2024]. Verfügbar unter: https://de.statista.com/statistik/daten/studie/4155/umfrage/entwicklung-der-werbeinvestitionen-in-deutschland-seit-2000/

Olson, Randal S.; Bartley, Nathan; Urbanowicz, Ryan J. und Moore, Jason H., 2016. Evaluation of a tree-based pipeline optimization tool for automating data science. In: *Proceedings of the genetic and evolutionary computation conference 2016*. New York, NY, USA: ACM. 2016. S. 485–492. Gecco '16. ISBN 978-1-4503-4206-3

Perlich, Claudia, 2011. Learning Curves in Machine Learning. In: Sammut, Claude und Webb, Geoffrey I. (Hrsg.), *Encyclopedia of Machine Learning*. Boston, MA: Springer US. S. 577–580. ISBN 978-0-387-30768-8

Pick, Doreén, 2016. Churn Management. *WiSt - Wirtschaftswissenschaftliches Studium*. 2016. Bd. 45, Nr. 2, S. 60–64. DOI 10.15358/0340-1650-2016-2-60

Prokhorenkova, Liudmila; Gusev, Gleb; Vorobev, Aleksandr; Dorogush, Anna Veronika und Gulin, Andrey, 2018. CatBoost: unbiased boosting with categorical features. In: Bengio, S.; Wallach, H.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N. und Garnett, R. (Hrsg.), *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, Inc. 2018. S. 6639–6649

Provost, Foster und Fawcett, Tom, 2013. *Data science for business*. 1st ed. Sebastopol, CA: O'Reilly Media. ISBN 978-1-4493-6132-7

Raschka, Sebastian, 2020. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. *arXiv:1811.12808 [cs, stat]* [online]. 10 November 2020. Nr. v3. [Zugriff am: 8 Dezember 2020]. Verfügbar unter: http://arxiv.org/abs/1811.12808

Reichheld, Frederick F., 1999. *The loyalty effect: the hidden force behind growth, profits and lasting value*. reprint. Boston, Mass: Harvard Business School Press. ISBN 978-0-87584-448-0

Reichheld, Frederick F. und Sasser Jr., W. Earl, 1990. Zero defections: Quality comes to services. *Harvard Business Review*. 9 Oktober 1990. Bd. 68, Nr. 5, S. 105–111 Business Source Elite

Rokach, Lior, 2019. *Ensemble learning: pattern classification using ensemble methods*. Second edition. New Jersey London Singapore Beijing Shanghai Hong Kong Taipei Chennai: World Scientific. Series in machine perception and artificial intelligence, vol. 85. ISBN 9789811201950

Schmittlein, David C. und Peterson, Robert A., 1994. Customer Base Analysis: An Industrial Purchase Process Application. *Marketing Science*. Februar 1994. Bd. 13, Nr. 1, S. 41–67. DOI 10.1287/mksc.13.1.41

Stauss, Bernd, 2000. Perspektivenwandel: Vom Produkt-Lebenszyklus zum Kundenbeziehungs-Lebenszyklus. *Thexis*. 7 März 2000. Bd. 17, Nr. 2, S. 15–18

Stauss, Bernd, 2011. Der Kundenbeziehungs-Lebenszyklus. In: Hippner, Hajo; Hubrich, Beate und Wilde, Klaus D. (Hrsg.), *Grundlagen des CRM*. Wiesbaden: Gabler. S. 319–341. ISBN 978-3-8349-2550-3

Tecklenburg, Tim, 2008. *Churn-Management im B2B-Kontext: eine empirische Analyse unter Berücksichtung von hierarchischen Kundenstrukturen und heterogenem Kundenverhalten*. 1. Aufl. Wiesbaden: Gabler. Gabler-Edition Wissenschaft. ISBN 978-3-8349-0866-7

Ting, Kai Ming, 2016. Confusion Matrix. In: Sammut, Claude und Webb, Geoffrey I. (Hrsg.), *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer US. S. 1–1. ISBN 978-1-4899-7502-7

Verbraken, Thomas; Verbeke, Wouter und Baesens, Bart, 2013. A Novel Profit Maximizing Metric for Measuring Classification Performance of Customer Churn Prediction Models. *IEEE Transactions on Knowledge and Data Engineering*. Mai 2013. Bd. 25, Nr. 5, S. 961–973. DOI 10.1109/TKDE.2012.50

Wang, Chi; Wu, Qingyun; Weimer, Markus und Zhu, Erkang, 2021. FLAML: A Fast and Lightweight AutoML Library. In: *MLSys*. 2021

WER, 2024. *Homepage WER GmbH* [online]. 2024. [Zugriff am: 23 August 2024]. Verfügbar unter: https://wer-gmbh.de

Wikipedia, 2024. Training, validation, and test data sets. *Wikipedia* [online]. [Zugriff am: 18 Februar 2024]. Verfügbar unter: https://en.wikipedia.org/w/index.php?title=Training,_validation,_and_test_data_sets&oldid=1199288619

Wirth, R. und Hipp, Jochen, 2000. CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*. Januar 2000

Wübben, Markus und Wangenheim, Florian v., 2008. Instant Customer Base Analysis: Managerial Heuristics Often “Get it Right”. *Journal of Marketing*. Mai 2008. Bd. 72, Nr. 3, S. 82–93. DOI 10.1509/jmkg.72.3.082

Hochschule Düsseldorf
University of Applied Sciences

Fachbereich
Wirtschaftswissenschaften
Faculty of
Business Studies

Düsseldorf Working Papers in Applied Management and Economics