Technische Hochschule Rosenheim

Fakultät für Wirtschaftsingenieurwesen

Masterarbeit

Zur Erlangung des akademischen Grades

Master of Engineering (M. Eng.)

Entwicklung eines Lead-Scoring-Modells für einfache und fortgeschrittene Anwendungsfälle am Beispiel der Software Mautic

Firma: Leuchtfeuer Digital Marketing GmbH

Immengarten 16-18

30177 Hannover

Erstprüfer: Prof. Dr. Peter Kraus

Zweitprüfer: Prof. Dr. Robert Kuttler

Vorgelegt am: 30.03.2024

Erstellt von: Jonas Ludwig

Münchener Str. 16

83527 Haag

Matrikel-Nr.: 889172

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel verwendet sowie wörtliche und sinngemäße Zitate als solche gekennzeichnet habe. Diese Arbeit wurde bei keiner anderen Stelle für einen ähnlichen Zweck vorgelegt. Die Versicherung erstreckt sich auch auf graphische Darstellungen und auf beigefügte oder zugrunde gelegte Software.

Haag, 30.03.2024 — J. Ludwig

Ort, Datum — Jonas Ludwig

Kurzfassung

Im Rahmen des digitalen Marketings haben Unternehmen die Möglichkeit, online Leads zu generieren und in ihren Datenbanken zu speichern. Da nicht alle dieser Leads ein konkretes Kaufinteresse aufweisen, ist es erforderlich, diese zu priorisieren, um nur qualifizierten Leads weitere Ressourcen zuzuteilen. Eine Methode dazu ist das Lead Scoring. Traditionell erhält dabei jeder Lead einen Score, der auf dessen Eigenschaften und Verhalten basiert. Überschreitet der Score einen Schwellenwert, gilt der Kontakt als qualifiziert und wird an das Vertriebsteam übergeben. Neben dem traditionellen Ansatz existiert auch der prädiktive Ansatz, der Algorithmen des maschinellen Lernens verwendet, um Leads zu qualifizieren.

Ziel dieser Arbeit ist die Entwicklung eines Lead-Scoring-Modells für einfache und erweiterte Anwendungsfälle am Beispiel der Software Mautic. Dazu wird zunächst ein generisches Vorgehensmodell zur Entwicklung traditioneller und prädiktiver Lead-Scoring-Systeme erstellt, welches anschließend an die Software Mautic angepasst wird. Zudem wird die Implementierung der erweiterten Anwendungsfälle des Account Based Scorings und des produktbasierten Lead Scoring in Mautic untersucht.

Die Ergebnisse der Forschungsarbeit zeigen, dass die Funktionalitäten von Mautic ausreichen, um sowohl den traditionellen als auch den prädiktiven Ansatz des Lead Scorings umzusetzen. Auch die Implementierung eines produktbasierten Lead Scorings ist möglich. Hingegen reichen die Funktionalitäten der Software derzeit nicht aus, um ein System für das Account Based Scoring zu erstellen. Um ein datengetriebenes Lead Scoring betreiben zu können, ist es außerdem notwendig, eine externe Software zur Datenauswertung in den Prozess zu integrieren. Die im Rahmen dieser Arbeit durchgeführten Praxistests zeigen, dass der Ansatz des prädiktiven Lead Scorings akkuratere Vorhersagen zur Lead-Qualität ermöglicht als der traditionelle Ansatz.

Abstract

Digital marketing gives companies the ability to generate leads online and store them in their databases. Since not all of these leads necessarily show a clear interest in making a purchase, it is necessary to prioritize them so that only qualified leads are allocated further resources. One method to do this is lead scoring. Traditionally, each lead is assigned a score based on their characteristics and behavior. If the score exceeds a certain threshold, the contact is considered qualified and forwarded to the sales team. In addition to the traditional approach, there is also the predictive approach, which utilizes Machine-Learning-Algorithms to qualify leads.

The objective of this work is to develop a lead scoring model for both simple and advanced use cases, using the software Mautic. To achieve this, a generic model for developing both traditional and predictive lead scoring systems is created and then adapted to the Mautic software. In addition, the implementation of advanced use cases such as account-based scoring and product-based lead scoring in Mautic will be examined.

The research results indicate that the functionalities of Mautic are sufficient to implement both the traditional and predictive lead scoring approaches. The implementation of product-based lead scoring is also feasible. However, the current capabilities of Mautic are not sufficient to create an account-based scoring system. In addition, a data-driven lead scoring system requires the use of external data analysis software. Real-world tests reveal that the predictive lead scoring approach enables more accurate predictions about the quality of a lead than the traditional approach.

Inhaltsverzeichnis

Erklärung I

Kurzfassung II

Abstract III

Inhaltsverzeichnis IV

Abbildungsverzeichnis VI

Tabellenverzeichnis VIII

Abkürzungsverzeichnis IX

1 Einleitung 1

1.1 Problemstellung und Zielsetzung 1

1.2 Methodik 2

2 Grundlagen 4

2.1 Lead 4

2.2 Customer Journey 4

2.3 Sales Funnel 4

2.4 Lead Management 6

2.5 Marketing Automation 7

3 Einordnung des Begriffs Lead Scoring 9

3.1 Traditionelles Lead Scoring 9

3.2 Ziele des Lead Scorings 13

3.3 Herausforderungen im Lead Scoring 14

4 Erweiterte Ansätze und Anwendungsfälle des Lead Scorings 22

4.1 Prädiktives Lead Scoring 22

4.2 Produktbasiertes Scoring 30

4.3 Account Based Scoring 31

4.4 Lead Scoring ohne Vertriebsteam 32

4.5 Weitere Anwendungsfälle 33

5 Alternative und ergänzende Methoden zum Lead Scoring 33

5.1 RFM-Analyse 33

5.2 Produktempfehlungssysteme 35

6 Entwicklung eines generischen Vorgehensmodells zum Lead Scoring 37

6.1 Traditionelles Lead Scoring 38

6.2 Prädiktives Lead Scoring 44

7 Anwendung des Vorgehensmodells auf die Software Mautic 45
7.1 Einführung der Software Mautic 45
7.2 Traditionelles Lead Scoring in Mautic 49
7.3 Prädiktives Lead Scoring in Mautic 59
7.4 Produktbasiertes Lead Scoring in Mautic 62
7.5 Account Based Scoring im Mautic 62
8 Praktische Umsetzung und Ergebnisse des Lead Scorings in Mautic 63
8.1 Traditionelles Lead Scoring 63
8.2 Prädiktives Lead Scoring 73
9 Zusammenfassung und Ausblick 76
Literaturverzeichnis X
Anhang XVII
Anhang 1: Kampagne zum Generieren von Vertriebsfeedback XVII
Anhang 2: Python-Code zur Aufbereitung der Lead-Daten XX
Anhang 3: SQL-Statements zum Download der Daten konvertierter Leads aus Mautic XXII
Anhang 4: SQL-Statements zum Download der Daten nicht-konvertierter Leads aus Mautic XXIII
Anhang 5: Python-Code zur Analyse der Lead-Daten XXIV
Anhang 6: Python-Code zum Ermitteln und Anpassen des Schwellenwertes im Lead Scoring XXVI
Anhang 7: Ergebnisse des prädiktiven Lead Scorings mit dem Support-Vector-Machine-Algorithmus XXVIII
Anhang 8: Ergebnisse des prädiktiven Lead Scorings mit dem Decision-Tree-Algorithmus XXIX
Anhang 9: Ergebnisse des prädiktiven Lead Scorings mit dem Logistic-Regression-Algorithmus XXX
Anhang 10: Python-Code zum Entwickeln eines Machine-Learning-Modells für das Lead Scoring XXXI
Anhang 11: Python-Code zur Anwendung des Machine-Learning-Modells XXXIII

Abbildungsverzeichnis

Abbildung 1: Traditioneller Sales Funnel 5
Abbildung 2: Lead-Scoring-Matrix 13
Abbildung 3: Inhaltsbasierte Produktempfehlungssysteme 35
Abbildung 4: Kollaborative Filtersysteme 36
Abbildung 5: Unterschiede zwischen dem traditionellen und dem prädiktiven Vorgehensmodell zur Erstellung eines Lead-Scoring-Systems 37
Abbildung 6: Benutzeroberfläche der Software Mautic 46
Abbildung 7: Point Groups in Mautic 51
Abbildung 8: Point Actions in Mautic 51
Abbildung 9: Kampagnenaktion "Adjust contact points" 52
Abbildung 10: Kampagne zum Errechnen des expliziten Scores 53
Abbildung 11: Kampagne zum Implementieren eines Verfallsmodells 54
Abbildung 12: Kampagne zum Zurücksetzen negativer impliziter Scores 55
Abbildung 13: Kampagne zur Übergabe von Leads 56
Abbildung 14: Kampagne zur Feedbackauswertung 57
Abbildung 15: Kampagne zum Reaktivieren von Leads 58
Abbildung 16: Import der Vorhersagen eines prädiktiven Lead-Scoring-Modells in Mautic 61
Abbildung 17: Kampagne zum Errechnen des impliziten Scores im Account Based Scoring 63
Abbildung 18: Korrelationsanalyse mit dem Lead-Status 68
Abbildung 19: Vergleich der Schwellenwerte konvertierter und nicht-konvertierter Leads 70
Abbildung 20: Konfusionsmatrix des vorläufigen traditionellen Lead-Scoring-Systems 71
Abbildung 21: Konfusionsmatrix des finalen traditionellen Lead-Scoring-Systems 72
Abbildung 22: Konfusionsmatrix zum prädiktiven Lead Scoring mit dem Random-Forest-Algorithmus 74
Abbildung 23: Kampagne zum Generieren von Vertriebsfeedback XVII
Abbildung 24: E-Mail zum Generieren von Vertriebsfeedback XVIII
Abbildung 25: JavaScript-Code zum Abschicken des Vertriebsfeedbacks XIX
Abbildung 26: Python-Code zur Aufbereitung der Lead-Daten XXI
Abbildung 27: SQL-Statement zum Download der Daten konvertierter Leads aus Mautic XXII
Abbildung 28: SQL-Statement zum Download der Daten nicht-konvertierter Leads aus Mautic XXIII
Abbildung 29: Python-Code zur Analyse der Lead-Daten XXV
Abbildung 30: Python-Code zum Ermitteln und Anpassen des Schwellenwertes im Lead Scoring XXVII
Abbildung 31: Ergebnisse des prädiktiven Lead Scorings mit dem Support-Vector-Machine-Algorithmus XXVIII

Abbildung 32: Ergebnisse des prädiktiven Lead Scorings mit dem Decision-Tree-Algorithmus XXIX
Abbildung 33: Ergebnisse des prädiktiven Lead Scorings mit dem Logistic-Regression-Algorithmus XXX
Abbildung 34: Python-Code zum Entwickeln eines Machine-Learning-Modells für das Lead Scoring XXXII
Abbildung 35: Python-Code zur Anwendung des Machine-Learning-Modells XXXIII

Tabellenverzeichnis

Tabelle 1: Zusammenhang zwischen den Komponenten einer Marketing-Automationssoftware und dem Lead Management 9
Tabelle 2: Explizite Lead-Scoring-Parameter 10
Tabelle 3: Positive und negative implizite Lead-Scoring-Parameter 11
Tabelle 4: Implizite Scorecard im Lead Scoring 12
Tabelle 5: KPIs im Lead Scoring 21
Tabelle 6: Unterschiede im traditionellen und prädiktiven Lead Scoring 23
Tabelle 7: Konfusionsmatrix 29
Tabelle 8: Unterschiede im B2C- und B2B-Markt 32
Tabelle 9: RFM-Tabelle zur Punktvergabe pro Einkauf 34
Tabelle 10: Errechnung des RFM-Scores 34
Tabelle 11: DataFrame konvertierter Leads 64
Tabelle 12: Leadübersicht DataFrame 66
Tabelle 13: Statistiktabelle für die Datenanalyse 67
Tabelle 14: Punktesystem im traditionellen Lead-Scoring-Praxisbeispiel 69
Tabelle 15: Tabelle zur Analyse der Score-Perzentile 71

Abkürzungsverzeichnis

ABS	Account Based Scoring
B2B	Business-to-Business
B2C	Business-to-Consumer
CRISP-DM	Cross-Industry Standard Process for Data Mining
CRM	Customer Relationship Management
CTA	Call-to-Action
DF	DataFrame
ERP	Enterprise Resource Planning
FN	False Negative
FP	False Positive
KPI	Key Performance Indicator
MAS	Marketing-Automationssoftware
MQL	Marketing Qualified Lead
SAL	Sales Accepted Lead
SLA	Service Level Agreement
TN	True Negative
TP	True Positive

1 Einleitung

1.1 Problemstellung und Zielsetzung

Im Kontext des digitalen Marketings haben Unternehmen die Möglichkeit, online Leads zu generieren und in ihren Datenbanken zu speichern. Nicht alle dieser Leads zeigen jedoch ein konkretes Kaufinteresse. Folglich gilt es im Rahmen der effizienten Nutzung der Vertriebsressourcen, Leads zunächst zu qualifizieren, um nur hochwertige Leads weiter zu bearbeiten, beispielsweise durch die Übergabe an das Vertriebsteam (vgl. Duncan und Eklan 2015: 1752). Eine Möglichkeit hierzu ist das Lead Scoring. Beim traditionellen Lead-Scoring-Ansatz erhält jeder Lead einen Score. Diesem werden auf Basis der Eigenschaften und des Verhaltens des Leads Punkte hinzugefügt oder abgezogen. Überschreitet der Score eines Kontakts nun einen vordefinierten Schwellenwert, so gilt dieser als qualifiziert und wird an das Vertriebsteam übergeben oder mit speziellen Marketingmaßnahmen für qualifizierte Leads bearbeitet (vgl. Schüller und Schuster 2022: 162–168). Neben dem traditionellen Ansatz des Lead Scorings, welcher Punktesysteme verwendet, existiert in der Praxis auch der prädiktive Ansatz. Dieser nutzt Machine-Learning-Algorithmen und Data-Mining-Modelle, um Leads für die Weitergabe an die Vertriebsabteilung zu qualifizieren (vgl. Wu et al. 2023: 1).

Zum Zeitpunkt der Erstellung dieser Arbeit liegen mehrere wissenschaftliche Forschungsarbeiten vor, welche die praktische Anwendung des traditionellen Lead Scorings (vgl. Monat 2011: 178–194; Naveen und Hariharanath 2021: 1302–1309; Verma et al. 2016: 220–237) sowie verschiedener Machine-Learning-Algorithmen im Rahmen des prädiktiven Lead Scorings untersuchen (vgl. Duncan und Eklan 2015: 1751–1758; Buckinx und van den Poel 2005: 252–268; Bohanec et al. 2015: 338–352; D'Haen et al. 2016: 69–78; Espadinha-Cruz et al. 2021: 1–14; Bohanec et al. 2017: 416–428; D'Haen und van den Poel 2013: 544–551; Kazemi und Babaei 2011: 37–45; Gokhale und Joshi 2018: 279–291; Jadli et al. 2022: 433–443; Nygard und Mezei 2020: 1439–1448; Kim und Street 2004: 215–228). Einige dieser Arbeiten, darunter Bohanec et al. (vgl. 2017: 416–428), Gokhale und Joshi (vgl. 2018: 279–291), Jadli et al. (vgl. 2022: 433–443) und Nygard und Mezei (vgl. 2020: 1439–1448) setzen sich dabei unter anderem mit der Wahl des besten Machine-Learning-Algorithmus im prädiktiven Lead Scoring auseinander. Wu et al. (vgl. 2023: 1–30) untersuchen außerdem den aktuellen Stand des Lead Scorings und dessen Einfluss auf das Vertriebsergebnis. Hierzu werden die Ergebnisse aus 44 Forschungsarbeiten, in welchen unterschiedliche traditionelle und prädiktive Lead-Scoring-Systeme entwickelt werden, verglichen und die Ergebnisse der einzelnen Arbeiten evaluiert.

Obwohl die vorliegenden Forschungsergebnisse darauf hindeuten, dass das prädiktive Lead Scoring zu besseren Ergebnissen führt als das traditionelle Lead Scoring, gibt es zum Zeitpunkt der Erstellung dieser Arbeit keine Forschung, die die Ergebnisse beider Ansätze anhand desselben Praxisbeispiels vergleicht. Dieser Umstand erschwert die Ableitung von Erkenntnissen über die Unterschiede in der Leistungsfähigkeit der beiden Ansätze. Darüber hinaus gibt es bisher keine Forschungsarbeit, welche die Umsetzung der verschiedenen Ansätze und Anwendungsfälle des Lead Scorings am Beispiel der Software Mautic untersucht. In der gegenwärtigen Situation mangelt es daher an einem methodisch fundierten Leitfaden für Nutzer, um ein effektives Lead-Scoring-System zu etablieren. Aufgrund fehlender Forschungsarbeiten zum Lead Scoring in Mautic ist zudem unklar, inwieweit die Funktionalitäten der Software ausreichen, um ein effektives Lead Scoring zu ermöglichen.

Angesichts der identifizierten Forschungslücken zielt diese Forschungsarbeit darauf ab, ein Lead-Scoring-Modell für einfache und fortgeschrittene Anwendungsfälle am Beispiel der Software Mautic zu entwickeln. Dieses Modell wird an einem praktischen Beispiel getestet, um dessen Leistungsfähigkeit zu validieren. Damit wird es den Anwendern der Software ermöglicht, deren Lead-Scoring-System an einem praxiserprobten Leitfaden auszurichten. Darüber hinaus wird die Forschungslücke geschlossen, dass die Ergebnisse des traditionellen und des prädiktiven Lead Scorings in der wissenschaftlichen Forschung noch nicht am gleichen Anwendungsbeispiel verglichen wurden. Ein weiteres Ziel dieser Arbeit ist es, die Funktionalitäten von Mautic im Bereich des Lead Scorings zu evaluieren. Basierend auf dieser Bewertung werden Empfehlungen für die Weiterentwicklung der Software gegeben, um so den Lead-Scoring-Prozess effektiver zu unterstützen und den Wert der Software aus Anwendersicht zu erhöhen.

1.2 Methodik

Zu Beginn dieser Forschungsarbeit wird eine umfassende Literaturrecherche durchgeführt, um mit dem Lead Scoring verbundene Begriffe zu spezifizieren. Anschließend wird der traditionelle Ansatz des Lead Scorings mit dessen Zielen und Herausforderungen behandelt. Zum Abschluss des theoretischen Teils werden erweiterte Ansätze und Anwendungsfälle des Lead Scorings beschrieben und alternative beziehungsweise ergänzende Methoden zum Lead Scoring betrachtet.

Im praktischen Teil werden auf Basis der zuvor identifizierten Herausforderungen generische Vorgehensmodelle zur Entwicklung robuster traditioneller und prädiktiver Lead-Scoring-Systeme erstellt. Anschließend wird die Software Mautic vorgestellt und die zuvor entwickelten Vorgehensmodelle auf diese Software angewandt. Darüber hinaus werden erweiterten Anwendungsfälle des produktbasierten und accountbasierten Lead Scorings in

Mautic untersucht. Anschließend werden das traditionelle und das prädiktive Lead Scoring anhand eines realen Praxisbeispiels in Mautic implementiert. Abschließend werden die Ergebnisse der praktischen Untersuchung ausgewertet, interpretiert und mit den Erkenntnissen der Literaturrecherche verglichen. Der Fokus liegt hierbei insbesondere auf der Bewertung der einzelnen Anwendungsfälle und den Unterschieden in den Ergebnissen des traditionellen und des prädiktiven Lead-Scoring-Ansatzes. Des Weiteren werden Möglichkeiten zur zielgerichteten Verbesserung der Software Mautic ausgearbeitet, um eine zielgerichtetere Implementierung des Lead Scoring zu ermöglichen.

2 Grundlagen

2.1 Lead

Um den Begriff "Lead Scoring" im Marketingkontext angemessen erläutern zu können, ist eine genaue Definition des Begriffs "Lead" notwendig. In dieser Arbeit wird auf die Definition von Todor (vgl. 2016: 90) Bezug genommen. Dieser definiert Leads als potenzielle Interessenten oder bereits bestehende Kunden, die Interesse an einem neuen Produkt oder einer neuen Dienstleistung zeigen.

2.2 Customer Journey

Leads durchlaufen auf ihrem Weg zum Kauf und darüber hinaus die sogenannte Customer Journey. Die Customer Journey umfasst die Gesamtheit aller Interaktionen eines Leads mit einem Unternehmen über alle Touchpoints hinweg. Bei den Touchpoints handelt es sich um Schnittstellen, die es den Leads ermöglichen, mit einem Unternehmen in Kontakt zu treten. Das grundlegende Ziel innerhalb der Customer Journey ist, dem Interessenten oder Kunden die richtigen Informationen zur richtigen Zeit und in der gewünschten Form zur Verfügung zu stellen und somit den Kaufprozess zu unterstützen. Die strategische Gestaltung der Customer Journey ist daher essenziell, um mehr Umsatz zu generieren (vgl. Hannig 2021: 246; Koerner 2021: 67–69).

2.3 Sales Funnel

Ein Sales Funnel dient zur Visualisierung des Qualifizierungsprozesses eines Leads innerhalb der Customer Journey (vgl. Hannig 2021: 247). In dieser Arbeit wird ein traditionelles Sales-Funnel-Modell verwendet, welches in Abbildung 1 dargestellt ist und Leads in vier Qualifizierungsstufen unterteilt: Lead, Marketing Qualified Lead (MQL), Sales Accepted Lead (SAL) und Sales Qualified Lead (SQL). Nachdem ein Lead diese vier Stufen durchlaufen hat, wird darauf abgezielt, einen Abschluss beziehungsweise "Close" zu erreichen. In den folgenden Abschnitten werden die einzelnen Phasen des Sales Funnels erläutert.

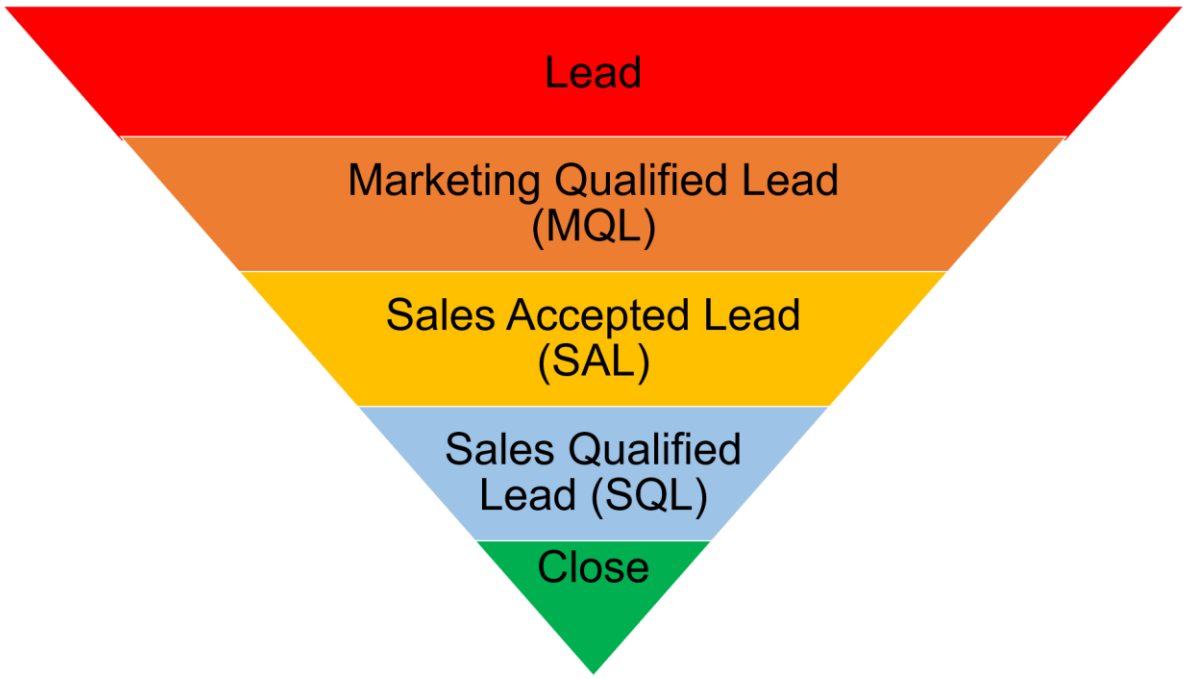

Abbildung 1: Traditioneller Sales Funnel
Quelle: In Anlehnung an Schüller und Schuster 2022: 148 und Schuster 2022: 89

Lead

Leads sind potenzielle Interessenten oder bestehende Kunden, die Interesse an einem neuen Produkt oder einer neuen Dienstleistung zeigen (siehe Kapitel 2.1).

Marketing Qualified Lead

Diese Leads werden vom Marketingteam so lange betreut und mit Marketingmaßnahmen bearbeitet, bis sie als verkaufsreif eingestuft werden. Ist dies der Fall, erhalten sie den Status MQL und werden an das Vertriebsteam übergeben (vgl. Philipp 2021: 207–208).

Sales Accepted Lead

Im Anschluss an die Übergabe werden die Leads von den Vertriebsmitarbeitern genauer überprüft und entweder akzeptiert oder abgelehnt. Entspricht ein Lead den von Marketing und Vertrieb gemeinsam festgelegten Parametern und wird dieser vom Vertrieb akzeptiert, ändert sich dessen Status in SAL (vgl. Schüller und Schuster 2022: 149).

Sales Qualified Lead

Die Kontakte durchlaufen nun einen Vertriebsprozess, um diese bis zur letzten Phase im Sales Funnel, dem Abschluss, zu führen (vgl. Schüller und Schuster 2022: 149). Dabei geht es insbesondere um die Qualifizierung der Leads aus Sicht des Vertriebs. Werden diese vom Vertriebsteam als Lead mit hohem Potenzial eingestuft, so erhalten sie den Status Sales Qualified Lead. Um einen Kontakt als SQL zu klassifizieren, müssen die folgenden BANT-Kriterien erfüllt sein (vgl. Hannig 2020: 218–219):

- Budget (B): Es ist ein ausreichendes Budget vorhanden.

- Authority (A): Der Kontakt ist autorisiert, die Kaufentscheidung zu treffen oder hat Einfluss auf die Kaufentscheidung.
- Need (N): Der Kontakt hat einen Bedarf, der zum Angebot des Vertriebs passt.
- Time (T): Der Kontakt hat die Absicht, in absehbarer Zeit zu kaufen.

2.4 Lead Management

Das Lead Management ist ein systematisierter Prozess, der darauf abzielt, Leads durch den gesamten Sales Funnel bis hin zum Vertriebsteam oder zum Abschluss zu bringen (vgl. Koerner 2021: 72). Das Lead Management lässt sich in folgende Bereiche unterteilen (vgl. Jörvinen und Taiminen 2016: 170–172):

1. Leadgenerierung und -identifizierung: Der erste Schritt im Lead Management ist die Leadgenerierung bzw. -identifizierung. Hierbei gibt ein Lead seine Kontaktdaten an das Unternehmen weiter, z. B. durch das Ausfüllen eines Formulars. Das Verhalten bestehender Kontakte kann anschließend durch verschiedene Methoden wie zum Beispiel das Tracking der IP-Adresse, Cookies oder Website-Logins erfasst werden.
2. Lead Nurturing: Unter Lead Nurturing versteht man Marketingprozesse, die darauf abzielen, Leads in MQLs umzuwandeln. Dabei geht es darum, die Kontakte durch aussagekräftige und relevante Inhalte der Kaufentscheidung näherzubringen. Die Inhalte werden dabei auf Basis der Profilinformationen eines Kontakts wie Unternehmen, Branche oder Position und dessen Online-Verhalten personalisiert.
3. Leadqualifizierung: Da nicht alle interessierten Leads proaktiv das Verkaufsteam kontaktieren und nicht alle Leads das gleiche Kundenpotenzial haben, werden Systeme zur Leadqualifizierung eingesetzt. Diese nutzen gesammelte Informationen über Leads, um darauf basierend eine Priorisierung vorzunehmen. Durch die Vergabe von Punkten basierend auf Verhalten und Merkmalen werden Leads mit höheren Punktzahlen bevorzugt an den Vertrieb weitergeleitet.
4. Leadübergabe: Qualifizierte Leads werden automatisch an das Customer Relationship Management (CRM)-System transferiert und auf Lead-Warteschlangen verteilt, die eingehende Leads den geeigneten Vertriebsteams zuweisen. Diese Warteschlangen können beispielsweise nach geografischem Standort und Geschäftsbereich kategorisiert werden. Jedes Vertriebsteam ist dabei für die Abarbeitung mindestens einer Warteschlange zuständig.
5. Abschluss: Nach der Übergabe ergreift das Unternehmen Maßnahmen, um die qualifizierten Leads zum Abschluss zu bewegen. Die entsprechenden Informationen werden im CRM-System festgehalten, unabhängig davon, ob ein Kontakt zum Kunden wird oder nicht. In einem optimalen Szenario ermöglicht die systematische

Organisation die Nachverfolgung des gesamten Marketing- und Vertriebsprozesses, beginnend mit der Erfassung der Kontaktinformationen bis hin zum Kauf.

2.5 Marketing Automation

Marketing Automation bezieht sich auf den Einsatz von Technologien und Softwarelösungen zur Automatisierung von Aufgaben im Lead- und Bestandskundenmanagement (vgl. Biegel 2009: 203; Schüller und Schuster 2022: 71; Heinzelbecker 2021b: 141–142). Dazu gehören Prozesse wie die Kundensegmentierung, die Datenintegration und das Kampagnenmanagement. Durch den effizienten Einsatz von Marketing-Automationssoftware (MAS) können relevantere Inhalte versendet und Leads effektiver in zahlende Kunden konvertiert werden. Folglich beschleunigt sich der Fortschritt der Leads durch die verschiedenen Stufen des Sales Funnels. Gleichzeitig sinken der Personalaufwand und die damit verbundenen Kosten. Eine erfolgreiche Automatisierung erfordert jedoch strategische Planung, Anpassungsfähigkeit und eine durchdachte Ausrichtung der Kampagnen und der Customer Journey, um die Zielgruppe zum richtigen Zeitpunkt mit relevanten Angeboten zu erreichen (vgl. Bagshaw 2015: 84–85). Beim Untersuchen der möglichen Funktionen und Bestandteile einer Marketing-Automationssoftware wurden unter anderem folgenden ermittelt:

Schnittstellen und CRM-Integration

Eine Grundlage für eine effiziente Marketing Automation ist die Anbindung des MAS an andere Systeme, um sicherzustellen, dass das MAS auf Daten aus diesen Systemen zugreifen und gleichzeitig Daten in diese Systeme übertragen kann. Insbesondere die Anbindung des MAS an das CRM-System ist notwendig, damit Marketing und Vertrieb effizient zusammenarbeiten können, zum Beispiel bei der Übergabe von MQLs an den Vertrieb (vgl. Griebsch 2021).

Segmentierung und dynamischer Content

Im Zuge der Segmentierung werden Kontakte aufgrund ihrer Eigenschaften oder ihres Verhaltens in verschiedene Zielgruppen oder Kundengruppen, auch Segmente genannt, unterteilt. Die einzelnen Segmente erhalten dann personalisierte Inhalte oder Werbebotschaften. Durch die Segmentierung können auch dynamische Inhalte für die einzelnen Kontakte generiert werden. Das bedeutet, dass auf Websites und anderen Plattformen unterschiedliche Inhalte angezeigt werden, je nachdem, in welchem Segment sich ein Kontakt befindet (vgl. Griebsch 2021; Teiu 2021: 330–331; DMEXCO o. D.).

Kampagnen

Marketingkampagnen stellen sicher, dass Kontakte zum optimalen Zeitpunkt die passenden Marketinginhalte erhalten. In Kampagnen werden verschiedene Marketingaktionen

miteinander verknüpft und automatisch in einer vordefinierten Reihenfolge an die Leads verschickt. Die Inhalte orientieren sich dabei an den Interessen und dem Verhalten der potenziellen Kunden, um eine individuelle und relevante Ansprache zu gewährleisten (vgl. Teiu 2021: 330–331).

E-Mail-Marketing

Im Bereich des E-Mail-Marketings ist es notwendig, dass Leads zunächst im Rahmen des Opt-In-Verfahrens zustimmen, Inhalte eines Unternehmens per E-Mail zu erhalten. Nach dieser Zustimmung können die Leads im Rahmen des Lead Nurturings gezielt mit relevanten Inhalten per E-Mail versorgt werden. Mithilfe einer MAS können die versendeten E-Mails durch die Integration von Empfängerdaten wie dem Namen, dem Firmennamen und der Telefonnummer personalisiert werden, um die Kommunikation individueller zu gestalten (vgl. Teiu 2021: 330).

Landing Pages

Landing Pages sind Webseiten, die dazu dienen, Informationen zu präsentieren oder die Eingabe von Kontaktdaten in Form von Formularen zu ermöglichen (vgl. Teiu 2021: 331).

Formulare

Formulare bieten Leads die Möglichkeit, Informationen anzufordern, sich für Veranstaltungen anzumelden oder eine Kontaktaufnahme durch das Unternehmen zu erbitten. Füllt ein Lead ein Formular aus und sendet dieses ab, ist dies für das Marketing ein deutliches Zeichen dafür, dass der Kontakt an einem Produkt, einer Dienstleistung oder einer Veranstaltung interessiert ist (vgl. DMEXCO o. D.; Teiu 2021: 331).

Social Media Marketing

Über soziale Medien kann Lead Nurturing betrieben werden. Außerdem können dort Formulare platziert werden, über die Kontakte identifiziert oder zu Marketingkampagnen hinzugefügt werden (vgl. Teiu 2021: 331).

Scoring-System

Ein weiterer möglicher Bestandteil einer MAS ist ein Lead-Scoring-System. In diesem werden Kontakte aufgrund ihres Verhaltens und ihrer Eigenschaften bewertet und priorisiert (vgl. Teiu 2021: 331–332).

Analytics + Reporting

Im Rahmen der einzelnen Kampagnen und Marketingmaßnahmen werden Daten generiert. Diese Daten werden im MAS gespeichert, um daraus Berichte zu erstellen oder Muster zu analysieren. So erhält das Unternehmen ein besseres Verständnis seiner Leads und kann seine Marketingmaßnahmen für die Zukunft optimieren (vgl. Teiu 2021: 332).

In Tabelle 1 werden die zuvor ermittelten Bestandteile des Lead Managements den einzelnen Schritten im Lead Management zugeordnet:

Lead generation and identification	Lead nurturing	Lead qualification	Lead routing	Closing
• Landing pages • Forms • Social media marketing	• Campaigns • Email marketing • Social media marketing • Landing pages • Segmentation and dynamic content	• Scoring system	• CRM integration	• Reporting tool

Tabelle 1: Zusammenhang zwischen den Komponenten einer Marketing-Automationssoftware und dem Lead Management
Quelle: Eigene Darstellung

3 Einordnung des Begriffs Lead Scoring

In diesem Abschnitt wird auf den traditionellen Lead-Scoring-Ansatz, die Ziele des Lead Scorings und die Herausforderungen im Lead Scoring eingegangen.

3.1 Traditionelles Lead Scoring

Wie in Kapitel 2.4 beschrieben, ist das Lead Scoring ein Bestandteil des Lead Managements, welcher es ermöglicht, Leads durch vordefinierte Kriterien nach ihrer Vertriebsreife zu beurteilen. Im Rahmen des Lead Scoring erhalten Leads Punkte basierend auf ihren Eigenschaften und Handlungen. Sobald der Score eines Leads einen vordefinierten Schwellenwert erreich, wird dieser als MQL eingestuft und an das Vertriebsteam weitergegeben (vgl. Hannig 2020: 217).

Implizite und explizites Lead Scoring

Bei der Auswahl der für das Lead Scoring relevanten Parameter wird zwischen impliziten und expliziten Parametern unterschieden. Das explizite Scoring konzentriert sich auf die Bewertung von Merkmalen der Interessenten. Dabei werden zum einen Informationen zur Person selbst und zum Unternehmen der Zielperson berücksichtigt. Durch das explizite Scoring wird ermittelt, wie gut ein Interessent zum Angebot des Unternehmens passt (vgl. Schüller und Schuster 2022: 164–165). Eine Sammlung möglicher expliziter Parameter, die aus den Werken von Schüller und Schuster (vgl. 2022: 164–165), Adobe (vgl. 2019: 18) und Patel (vgl. o. D.) hervorgehen, ist der Tabelle 2 zu entnehmen.

Explicit parameters for the lead	Explicit parameters for the lead's company
• Current position • Number of years in the current position • Department • Interests • Management affiliation (Yes/No) • Role in the decision-making process or buying center • Type of email address (company address or Gmail, Yahoo, etc.) • Lead source • Products already bought	• Number of employees • Revenue • Industry affiliation • Country • Budget availability • Time of the planned realisation of the solution • Year of establishment • Organizational structure • Website traffic • Financial condition of the company • Revenue growth

Tabelle 2: Explizite Lead-Scoring-Parameter
Quelle: Eigene Darstellung

Es ist zu beachten, dass sich explizite Parameter auch negativ auf den Score auswirken können. Wenn zum Beispiel ein Kontakt eine Position besetzt oder in einer Branche beschäftigt ist, die nicht zum Unternehmen passt, können Punkte abgezogen werden (vgl. Adobe 2019: 22).

Das implizite beziehungsweise handlungsbasierte Scoring konzentriert sich im Gegenzug zum expliziten Scoring auf die Bewertung des Verhaltens von Interessenten. Auch wenn ein Kontakt im expliziten Scoring eine hohe Bewertung erhält, sollte dieser nicht automatisch an die Vertriebsabteilung übergeben werden. Denn der explizite Score sagt lediglich aus, dass das Profil des Leads zum Unternehmen passt, nicht aber, ob der Lead bereits vertriebsreif ist. Die Vertriebsreife wird erst dann erreicht, wenn der Lead aktiv durch Handlungen mit dem Unternehmen interagiert. Diese Handlungen werden mit Hilfe des impliziten Lead Scorings bewertet (vgl. Schüller und Schuster 2022: 166). Innerhalb des impliziten Lead Scorings ist zudem zu beachten, dass es negative implizite Parameter gibt. Diese deuten darauf hin, dass Leads ein geringes Interesse am Unternehmen haben. Daher werden für negative Parameter Punkte vom Lead Score abgezogen (vgl. Naveen und Hariharanath 2021: 1306). Beispiele für positive und negative implizite Parameter im Lead Scoring, welche der untersuchten Literatur entnommen wurden (vgl. Adobe 2019: 19–21; Schüller und Schuster 2022: 166; Naveen und Hariharanath 2021: 1306–1307; Braun 2021: 163), sind Tabelle 3 zu entnehmen.

Positive implicit parameters	Negative implicit parameters
• Newsletter registration • Form submission • Blog page visit • Price page visit • Product page visit • Number of website visit • Website dwell time • Participation in webinars • Requesting informational material • Downloading informational materials • Reading reviews • Participation in forum or blog • Clicking links in emails • Email opens • Response to telemarketing campaigns • Attending an event • Social media engagement • Interaction with customer service • Utilizing a free trial • Visiting a trade fair	• Joining a do-not-call list • Long periods of website inactivity • Newsletter unsubscribe request • Unopened emails • Visiting non-commercial pages, such as the career page, indicating no purchasing intent • Spam complaint

Tabelle 3: Positive und negative implizite Lead-Scoring-Parameter
Quelle: Eigene Darstellung

Festlegen der Bepunktung

Beim traditionellen Lead Scoring wird zunächst durch Experten mit Verständnis der Customer Journey festgelegt, welche impliziten und expliziten Parameter für das eigene Lead Scoring relevant sind. Anschließend wird sowohl für die impliziten als auch für die expliziten Parameter eine Scorecard festgelegt. Darin werden alle relevanten Parameter aufgelistet und mit Punktwerten versehen, die bei Erfüllung des Parameters zum impliziten

oder expliziten Score addiert oder subtrahiert werden (vgl. Duncan und Eklan 2015: 1752; Schüller und Schuster 2022: 164–166). Tabelle 4 ist ein Beispiel für eine implizite Scorecard zu entnehmen.

Activity	Points
Initial contact via requesting a content offer	30
Link-click in a lead nurturing email	10
Form submission	10
Website or product page visit	5
Response to content offer Whitepaper A	20
Response to content offer Whitepaper B	35
Response to content offer Whitepaper C	25
Participation in a webinar	40
Trade fair attendance	45
Newsletter subscription	10
…	

Tabelle 4: Implizite Scorecard im Lead Scoring
Quelle: In Anlehnung an Schüller und Schuster 2022: 166

Festlegen eines Schwellenwertes

Bevor das Scoring-System in der Praxis eingesetzt werden kann, müssen noch die richtigen Schwellenwerte für den impliziten und expliziten Score festgelegt werden. Dies sind die Punktwerte, ab denen die Leads an den Vertrieb übergeben werden. Gleichzeitig können die MQLs mithilfe einer Lead-Scoring-Matrix priorisiert werden. Dabei werden die einzelnen Scores je nach Höhe in verschiedene Kategorien mit Schwellenwerten eingeteilt. Beispielsweise kann der implizite Score je nach Höhe in die Kategorien 1, 2 und 3 und der explizite Score in die Kategorien A, B, C und D eingeteilt werden (vgl. Naveen und Hariharanath 2021: 1306; Schüller und Schuster 2022: 167). Dementsprechend können die Leads anschließend in der Reihenfolge ihrer Priorität abgearbeitet werden, wobei beispielsweise die Leads der Kategorie A1 aus Abbildung 2 die höchste Priorität haben.

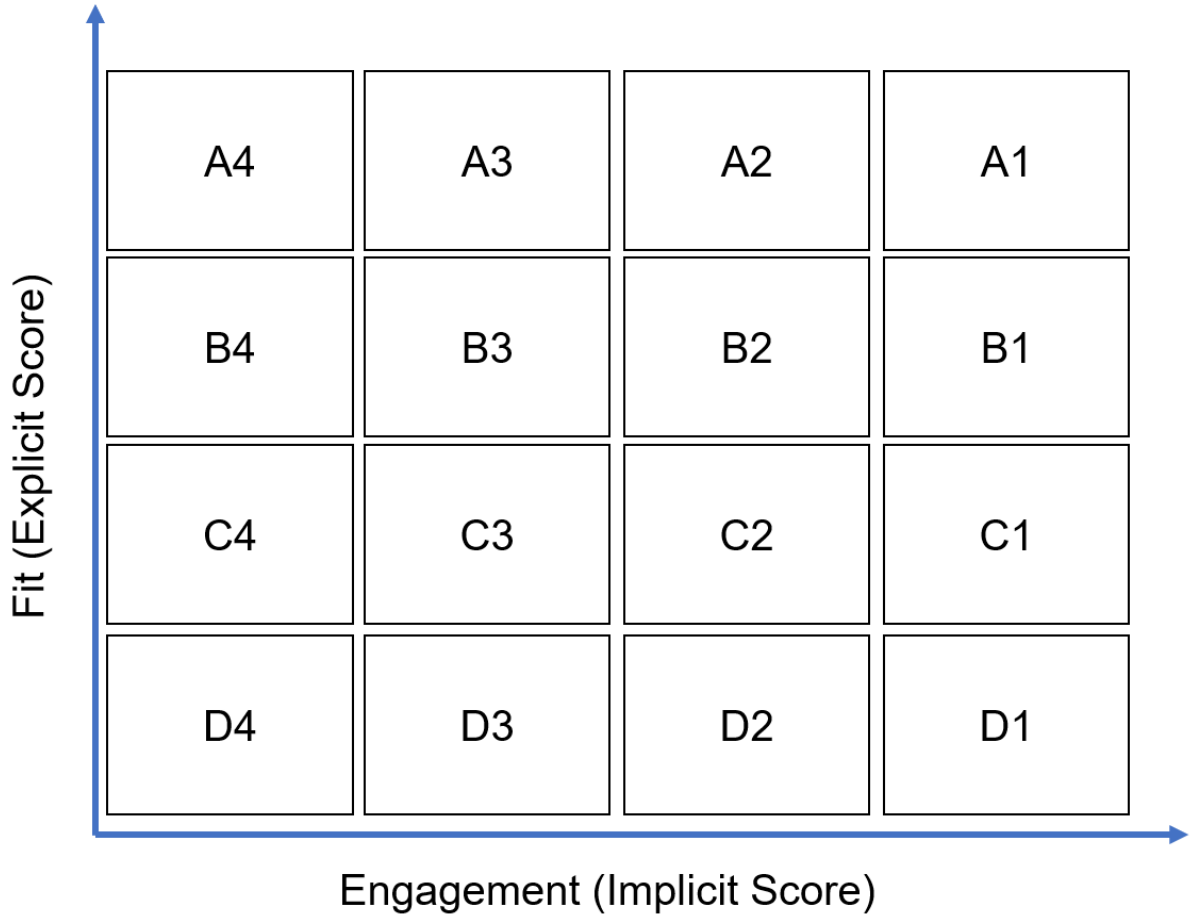

Abbildung 2: Lead-Scoring-Matrix
Quelle: In Anlehnung an Naveen und Hariharanath 2021: 1306

3.2 Ziele des Lead Scorings

Um das Lead Scoring näher einordnen zu können, wird im Folgenden auf die positiven Effekte des Lead Scorings eingegangen. Hierbei wurden unter anderem drei Hauptziele ermittelt.

Umsatzsteigerung

Im Marketing- und Vertriebsprozess kann es vorkommen, dass Leads im Verkaufsprozess kontaktiert werden, bevor diese reif für eine Kaufentscheidung sind. Dies führt nicht nur zu Frustration bei den Leads, sondern kann auch zum Verlust potenzieller Kunden führen. Das Lead Scoring ermöglicht es, reife und kaufbereite Leads zu identifizieren und zu priorisieren (vgl. Schüller und Schuster 2022: 147). In der Folge werden Leads erst dann an die Vertriebsmitarbeiter übergeben, wenn sie kaufbereit sind. Durch dieses verbesserte Timing werden somit mehr Leads konvertiert, wodurch der Umsatz steigt.

Verbessern der Zusammenarbeit zwischen Marketing und Vertrieb

Die Zusammenarbeit zwischen Marketing und Vertrieb ist entscheidend für die erfolgreiche Umwandlung von Leads in zahlende Kunden. Werden Leads zu früh übergeben, kann dies

zu Konflikten und Ineffizienzen in der Zusammenarbeit der beiden Abteilungen führen. Einerseits hat der Vertrieb in diesem Fall kein Vertrauen in die Qualität der übergebenen Leads und lehnt diese möglicherweise vorschnell ab, andererseits hat das Marketing das Gefühl, dass die Leads vom Vertrieb nicht ordnungsgemäß weiterverfolgt werden. Durch die Implementierung eines transparenten Lead-Scoring-Systems können diese Konflikte vermieden werden, da die Kriterien zur Qualifizierung und Priorisierung von Leads klar definiert sind (vgl. Hannig 2020: 218; Schüller und Schuster 2022: 147). Dadurch werden die Zusammenarbeit und das Verhältnis zwischen Marketing und Vertrieb verbessert.

Kosteneinsparung und Effizienzsteigerung

Ein zentrales Ziel des Lead Scorings ist die Rationalisierung des Vertriebsprozesses, um die Zeit- und Ressourceneffizienz zu maximieren. Wenn das Marketing zu viele Leads an den Vertrieb weiterleitet, können diese nicht mehr bearbeitet werden (vgl. Lontzek 2022; Duncan und Eklan 2015: 1752). Das Lead Scoring löst dieses Problem, da es sicherstellt, dass nur die relevantesten und qualifiziertesten Leads an den Vertrieb übergeben werden (vgl. Michiels 2008: 4). Dies ermöglicht eine gezieltere und effizientere Bearbeitung, da nur die Kontakte an die Vertriebsabteilung weitergeleitet werden, die tatsächlich an einem Kauf interessiert sind. Auch wenn keine Kapazitätsengpässe bestehen, kann Lead Scoring sinnvoll sein. Da uninteressante Kontakte automatisch herausgefiltert werden, kann der Vertrieb mehr Zeit und Ressourcen in vielversprechende Leads investieren. Dadurch steigen nicht nur die Chancen auf erfolgreichere Verkaufsabschlüsse, sondern auch die Kosten für die Leadbearbeitung sinken. Dies führt zu einer insgesamt ressourcenschonenderen Arbeitsweise im Vertrieb und steigert die Effizienz des gesamten Prozesses (vgl. Auerochs 2021).

3.3 Herausforderungen im Lead Scoring

Bei der Analyse der Literatur zum Thema Lead Scoring wurden mehrere Herausforderungen identifiziert, die sich auf den Erfolg von Lead-Scoring-Projekten auswirken können. Diese werden im Folgenden dargelegt.

Sicherstellen der Vollständigkeit und Qualität der Daten

Wie bereits im aktuellen Kapitel beschrieben, basiert das Lead Scoring auf der Erhebung, Speicherung und Nutzung personenbezogener Daten. Darum wird das Lead Scoring erschwert, wenn die benötigten Daten nicht in der erforderlichen Menge und Qualität zur Verfügung stehen (vgl. Monat 2011: 188; Lontzek 2022). Laut der Studie "Benchmarking Marketing Automation: The Shift Toward Next Generation Lead Scoring & Segmentation" von Decision Tree Labs ist aus Sicht der befragten Marketingexperten einer der beiden Hauptgründe für das Scheitern von Lead-Scoring-Projekten, dass unvollständige oder

inkonsistente Daten über Leads vorliegen (vgl. Lattice 2014: 6). Der Mangel an für das Lead Scoring relevanten Daten kann durch mehrere Gründe bedingt sein.

Einer davon sind Einschränkungen im Lead-Tracking. Um die Customer Journey eines Leads erfolgreich zu analysieren und somit verhaltensbasierte Daten für das Lead Scoring zu erheben, sind Tracking-Methoden notwendig. Die gängigste Tracking-Methode ist der Einsatz von Cookies (vgl. Flocke und Holland 2014: 216). Cookies sind kleine Textdateien, die beim Besuch einer Website auf dem Gerät des Besuchers gespeichert werden. Sie helfen dabei, die einzelnen Besucher eindeutig zu identifizieren und Informationen über diese zu speichern. Die Cookies werden dabei direkt im Browser des Nutzers gespeichert (vgl. Gradow und Greiner 2021: 5–6). In der gesichteten Literatur wurden einige Einschränkungen des Cookie Trackings identifiziert, die sich negativ auf die Beschaffung von aktionsbezogenen Lead-Daten auswirken können:

- Notwendigkeit einer Einwilligung: Sofern ein Cookie nicht zwingend erforderlich ist, um die Funktionalität einer Website zu gewährleisten, müssen Nutzerinnen und Nutzer nach der Datenschutz-Grundverordnung aktiv in die Verwendung von Cookies einwilligen. Erfolgt keine Einwilligung, muss den Nutzern dennoch die Website im vollen Umfang nutzen zu können (vgl. Gradow und Greiner 2021: 10–12). Von Nutzern, die der Verwendung von Cookies nicht zugestimmt haben, dürfen daher keine Daten für implizites Lead Scoring erhoben werden.
- Löschen von Cookies: Neben der Notwendigkeit, dass Nutzer Cookies aktiv zustimmen müssen, können Nutzer Cookies jederzeit innerhalb des Browsers löschen (vgl. Flocke und Holland 2014: 218), beispielsweise durch das Löschen des Browserverlaufs oder das Leeren des Cache. Darüber hinaus bieten Browser wie Safari, Google Chrome oder Mozilla Firefox die Möglichkeit, das Speichern von Cookies über die Einstellungen komplett zu unterbinden (vgl. Mozilla o. D.; Apple o. D.; Google o. D.). Innerhalb von Safari werden Cookies sogar standardmäßig nach sieben Tagen automatisch gelöscht, auch wenn die Nutzer der Speicherung eines Cookies auf einer Webseite zugestimmt haben (vgl. MM-Redaktion 2021). Aktionen, die Nutzer nach dem Löschen von Cookies durchführen, können somit nicht mehr eindeutig zugeordnet und für das Lead Scoring verwendet werden.
- Verwendung unterschiedlicher Browser und Geräte: Wie bereits beschrieben, werden Cookies im Browser eines Endgerätes gespeichert. Daher können Nutzer nicht identifiziert werden, wenn sie mit einem anderen Browser oder Gerät arbeiten. Füllt ein Interessent beispielsweise am Arbeitsplatz ein Formular aus und ruft die Website anschließend zu Hause am eigenen Rechner auf, werden die neuen Daten nicht dem zuvor generierten Leadprofil hinzugefügt (vgl. Adobe 2019: 39).

Neben dem Tracking von Website-Aktivitäten gibt es auch Probleme beim Tracking von E-Mails. Beim E-Mail-Tracking wird ein kleines Bild, auch Tracking-Pixel genannt, in den Text der E-Mail eingebettet. Anhand dieses Pixels kann dann festgestellt werden, ob eine E-Mail geöffnet wurde (vgl. Hu et al. 2019: 366). Zum Pixel-Tracking in E-Mails nennt Adobe zwei Quellen, die für fehlerhaftes Tracking sorgen können (vgl. Adobe 2019: 39):

- E-Mail-Öffnung durch Bots: Einige E-Mail-Anbieter setzen heutzutage Bots ein, die E-Mails automatisch öffnen, um diese auf Spam zu prüfen. Dadurch wird fälschlicherweise erfasst, dass ein Kontakt eine E-Mail geöffnet hat.
- Geblocktes HTML: Außerdem haben E-Mail-Empfänger die Möglichkeit das Laden von HTML-Anteilen in Mails zu blockieren. In diesem Fall wird der Tracking-Pixel nicht geladen. Folglich wird die E-Mail-Öffnung im Lead Scoring nicht erfasst.

Es gibt mehrere alternative Lead-Tracking-Technologien, die anstelle von Cookies und Pixeln eingesetzt werden können. Zu diesen gehören beispielsweise das Tracking über einen virtuellen Fingerabdruck, die IP-Adresse oder über die Anmeldung auf der Website. Diese alternativen Tracking-Methoden lösen einige der Probleme, die mit der Verwendung von Cookies verbunden sind. Allerdings weisen auch die alternativen Technologien Schwachstellen auf und garantieren daher kein lückenloses Lead-Tracking. In der Praxis wird im Rahmen des Customer Journey Trackings daher meist eine Kombination verschiedener Tracking-Methoden eingesetzt (vgl. Flocke und Holland 2014: 218–225; Jörvinen und Taiminen 2016: 170).

Neben den Einschränkungen beim Lead-Tracking stellt die Datensparsamkeit der Leads eine weitere Herausforderung für die Vollständigkeit und Qualität der Daten dar. Neue Leads sind oft nicht bereit, übermäßig viele persönliche Daten preiszugeben. Die Erhebung von zu vielen Daten in Formularen kann daher dazu führen, dass sich mehr Interessenten gegen die Angabe ihrer Daten entscheiden und somit die Conversion Rate, beispielsweise von Opt-In Formularen, sinkt (vgl. Schuster 2021: 110–111; Schüller und Schuster 2022: 154). Schüller und Schuster (vgl. 2022: 154–156) empfehlen daher den Einsatz von Techniken des Progressive Profiling. Progressive Profiling im Kontext von Lead Scoring bezeichnet eine schrittweise Datenerhebungsstrategie, bei der Unternehmen in mehreren Interaktionen schrittweise zusätzliche Informationen von Leads sammeln, anstatt alle erforderlichen Informationen in einem einzigen Datenerhebungsvorgang abzufragen. Diese Methode zielt darauf ab, die Konversionsrate zu erhöhen, indem der anfängliche Erfassungsaufwand minimiert wird, während im Laufe der Zeit ein umfassenderes Profil des Interessenten erstellt wird. Dieser schrittweise Ansatz verbessert die Qualität des Lead Scorings und ermöglicht eine genauere Qualifizierung potenzieller Kunden. Eine weitere Möglichkeit, welche im Lead Scoring Leitfaden von Adobe (vgl. 2019: 11) vorgestellt wird, ist das Zurückgreifen auf einen

Service, um fehlende Felder in der Lead Datenbank zu ergänzen. So kann eine vollständige Datengrundlage garantiert werden und gleichzeitig eine hohe Conversion Rate erreicht werden, da in Formularen nur wenige Informationen abgefragt werden.

Darüber hinaus stellt die Verwendung von Daten, die vor der Lead-Identifizierung erfasst wurden, eine weitere Herausforderung dar. Ein großer Teil der Customer Journey kann bereits stattfinden, bevor der Lead vom Unternehmen eindeutig identifiziert wurde, beispielsweise durch das Ausfüllen eines Opt-In Formulars. Grundsätzlich können diese Daten bereits vor der Erfassung des Leads anonym getrackt und nach der Identifizierung dem Lead-Profil zugeordnet werden. Diese Anreicherung des Leadprofils mit Daten vor der Registrierung ist in Deutschland allerdings nicht zulässig. Daher dürfen die Daten erst ab dem Opt-In verwendet werden (vgl. Schüller und Schuster 2022: 165–166).

Zusammenarbeit zwischen Marketing und Vertrieb

Die effiziente Durchführung von Lead Scoring erfordert eine enge Zusammenarbeit zwischen der Marketing- und der Vertriebsabteilung. Die Gestaltung dieser Zusammenarbeit kann sich als schwierig erweisen, insbesondere wenn beide Seiten unterschiedliche Vorstellungen vom Lead-Scoring-Prozess haben (vgl. Hannig 2020: 214). Dies kann dazu führen, dass die Vertriebsabteilung nicht mit den von der Marketingabteilung übergebenen Leads zufrieden ist (siehe Kapitel 3.2). Um eine reibungslose Zusammenarbeit zwischen Marketing und Vertrieb zu gewährleisten, ist es daher notwendig, dass Marketing und Vertrieb gemeinsam die Rahmenbedingungen für ihre Zusammenarbeit im Rahmen des Lead Scorings festlegen. Sind die Erwartungen beider Seiten aufeinander abgestimmt, kann die Zusammenarbeit effizienter gestaltet werden, da beide Seiten im Interesse der jeweils anderen handeln. Eine empfohlene Vorgehensweise zur schriftlichen Festlegung dieser Rahmenbedingungen ist die Verwendung eines Service Level Agreements (SLA). Ein SLA ist eine verbindliche Vereinbarung zwischen zwei Parteien über wiederkehrende Leistungen. In diesem Fall ist das SLA eine Art Vertrag zwischen der Vertriebs- und der Marketingabteilung, um eine optimale Betreuung der Leads zu gewährleisten. In diesem Dokument werden verschiedene Aspekte wie die einzuhaltenden Service Levels, der Leistungsumfang, die Art und Weise der Leistungserbringung und die zeitlichen Rahmenbedingungen geklärt. Da das SLA ein kritischer Faktor für das Lead Scoring ist, sollte das Dokument stets gepflegt und regelmäßig aktualisiert werden. Für ein SLA im Bereich Lead Scoring wird insbesondere empfohlen, die Antworten auf folgende Fragen zu dokumentieren (vgl. Schüller und Schuster 2022: 181–182):

- Welcher Zielgruppe wird für das Lead Scoring gewählt?
- Was ist ein idealer Lead, bzw. wie sieht das ideale Kundenprofil aus?
- Wie definiert sich ein MQL?

- Wie definiert sich ein SAL?
- Wie definiert sich ein SQL?
- Welche Daten werden zwischen dem MAS und dem Enterprise Resource Planning (ERP)- und CRM-System übertragen?
- Welche Stufen durchlaufen Interessenten im Rahmen der Customer Journey?
- Welche Daten werden im Lead-Nurturing-Prozess wie erfasst?
- Welche Methode zur Leadbewertung wird gewählt (zum Beispiel mit Punkten oder Noten in Form von Buchstaben)?
- Welche Schwellwerte werden für das Lead Scoring definiert?
- Welche Informationen zum Lead werden an den Vertrieb übergeben?
- Was geschieht nach der Leadübergabe?
- Was meldet das Vertriebsteam nach der Übergabe an die Marketingabteilung zurück?

Integration der Marketing-Automationssoftware

Eine weitere Voraussetzung für eine effektive Marketing Automation beziehungsweise für ein effektives Lead Scoring ist die Integration anderer Systeme, die entlang der Customer Journey eingesetzt werden. Durch die Integration können Daten automatisch in die MAS übernommen und für die Datenanalyse im Rahmen des Lead Scorings genutzt werden. Ist dies nicht der Fall, besteht die Gefahr, dass wichtige Daten für das Lead Scoring nicht zur Verfügung stehen (vgl. Rahimi 2020; Schoepf 2021: 284; Koerner 2021: 76; Griebsch 2021). Darüber hinaus können Daten an andere Systeme weitergegeben werden. Beispielsweise können für den Vertrieb relevante Daten an die CRM-Software übergeben werden (vgl. Michiels 2008: 16).

Identifizieren der Kaufwahrscheinlichkeitsfaktoren

Die Bedeutung der einzelnen impliziten und expliziten Parameter variiert je nach Unternehmen. Daher ist es eine Herausforderung, die ausschlaggebenden Kriterien, die sich auf die Leadqualität auswirken, zu identifizieren und zu gewichten (vgl. Monat 2011: 187). Laut der bereits unter in diesem Kaptiel genannten Studie "Benchmarking Marketing Automation: The Shift Toward Next Generation Lead Scoring & Segmentation", ist aus der Sicht der befragten Marketingexperten neben einer mangelnden Datengrundlage der zweite Hauptgrund für das Scheitern von Lead-Scoring-Vorhaben, ein mangelnder Einblick, welche impliziten und expliziten Daten tatsächlich Indizien für eine Kaufbereitschaft liefern (vgl. Lattice 2014: 9). Im traditionellen Lead Scoring werden Werte und Scores auf Basis von Experteneinschätzungen ermittelt. Hierbei handelt es sich jedoch nur um subjektive Annahmen und es ist nicht gewährleistet, dass diese tatsächlich der Realität entsprechen (vgl. Jadli et al. 2022: 434, 2022: 434; Duncan und Eklan 2015: 1752). Aus diesem Grund

wird empfohlen das Bewertungssystem auf Daten zu stützen (vgl. Bohanec et al. 2017). Daher empfiehlt es sich, statistische Methoden zu verwenden, um das Lead-Scoring-System auf Daten zu stützen.

Aussortieren irrelevanter Leads

In der Praxis sind nicht alle Kontakte in der MAS tatsächlich relevant. Bei den dort erfassten Kontakten kann nicht ausgeschlossen werden, dass es sich um Forschungsstudenten, Konkurrenten oder ähnliche Personen handelt, die ohne ein Kaufinteresse recherchieren. Daher muss verhindert werden, dass diese Leads das Vertriebsteam erreichen. Indizien hierfür können beispielsweise Begriffe wie "Student", "Professor" oder "arbeitslos" sein, die in Formularen als Position angegeben werden. In diesen Fällen muss der Status des Leads automatisch auf "unqualifiziert" gesetzt werden (vgl. Gooding 2022: 236).

Berücksichtigen nicht-linearer Effekte

Beim Lead Scoring mit einer Scorecard ist es nicht möglich, komplexe oder nicht-lineare Zusammenhänge zwischen den einzelnen Kriterien zu berücksichtigen. Besucht ein Lead beispielsweise mehrere Webinare, so erhöht sich dessen Score für jedes Webinar um die festgelegte Anzahl an Punkten. Es kann jedoch sein, dass die Kaufwahrscheinlichkeit eines Leads ab einer bestimmten Anzahl von Webinaren nur noch geringfügig ansteigt. Wenn beispielsweise die hochwertigsten Interessenten zwischen einem und drei Webinaren besucht haben, sind weitere Webinarbesuche nicht unbedingt ein Signal für eine höhere Kaufbereitschaft. Es kann sogar der Fall sein, dass der Besuch vieler Webinare ein negatives Signal ist. Dies könnte zum Beispiel auf das Verhalten von Studenten oder Konkurrenten hindeuten, die sich über das Angebot des Unternehmens informieren (vgl. Duncan und Eklan 2015: 1752).

Abhängigkeit von verhaltensbasierten Daten

Traditionelle Lead-Scoring-Modelle sind stark von impliziten Daten abhängig. Diese Daten können zwar ein guter Indikator für das Interesse von Leads am Angebot eines Unternehmens sein, sie können jedoch die frühzeitige Erkennung von hochwertigen Leads verhindern. Potenzielle Leads werden erst identifiziert, wenn diese ausreichend Aktionen durchgeführt haben, um einen hohen Score zu erreichen. Leads, die bereits von Anfang an kaufbereit sind, werden somit nicht identifiziert (vgl. Duncan und Eklan 2015: 1752). Als Lösung schlägt Gooding (vgl. 2022: 235) vor, im Rahmen der Leadbewertung einen sogenannten "Fast Track Path" zu entwickeln. Dabei werden Aktionen oder Kombinationen von Aktionen definiert, die darauf hindeuten, dass ein Lead von Anfang an kaufinteressiert ist. Führt ein Lead diese speziellen Aktionen aus, wird er direkt an das MQL und den Vertrieb weitergeleitet.

Berücksichtigung der zeitlichen Komponente

Eine Herausforderung im Lead Scoring ist die Berücksichtigung der zeitlichen Komponente. Werden vergangene Aktionen, wie der Besuch einer Produktwebseite vor zwei Jahren, genauso bewertet werden wie aktuelle, so kann dies zu Ineffizienzen im Lead Scoring führen. Der Grund dafür ist, dass vergangene Aktionen keinen Aufschluss darüber geben, ob der Interessent immer noch interessiert ist. Es wird daher empfohlen, vergangene Aktionen in Relation zur Zeit zu setzen, um eine genaue Bewertung der Leads zu ermöglichen (vgl. Kumar und Reinartz 2018: 151; Rahimi 2020).

Ermitteln des Schwellenwertes

Eine weitere Herausforderung ist das Festlegen des Schwellenwertes, bei dem Leads an das Vertriebsteam übergeben werden. Wird dieser zu niedrig angesetzt, werden zu viele unqualifizierte Leads übergeben. Wird dieser zu hoch angesetzt, besteht die Gefahr, dass qualifizierte Leads nicht identifiziert werden (vgl. Lontzek 2022). Monat (vgl. 2011: 191) und Adobe (vgl. 2019: 26) empfehlen hierbei, rückwirkend die Lead Scores von SALs zum Zeitpunkt der Übergabe zu berechnen und diese mit den Scores von Leads, die nicht übergeben wurden oder vom Vertrieb abgelehnt wurden zu vergleichen. Auf dieser Basis kann dann der Schwellenwert gewählt werden, der rückblickend die besten Ergebnisse erzielt hätte.

Übergabe der Leads an das richtige Vertriebsteam

Um ein zielführendes Lead Scoring zu gewährleisten, müssen die Leads an das richtige Vertriebsteam beziehungsweise die richtigen Vertriebsmitarbeiter übergeben werden. Dieser Zuweisungsprozess erfordert die Berücksichtigung der individuellen Kenntnisse und Fähigkeiten der Mitarbeiter, um sicherzustellen, dass die richtige Person den Kontakt bearbeitet (vgl. Schüller und Schuster 2022: 182). Um einen effizienten Prozess zu gewährleisten, empfiehlt es sich daher, den Zuordnungsschritt automatisiert in das Lead Scoring zu integrieren. Dabei kann eine Integration des CRM-Systems helfen, in dem die Leads anhand vordefinierter Regeln automatisiert dem entsprechenden Vertriebsteam oder den richtigen Vertriebsmitarbeitern zugeordnet werden (vgl. Jörvinen und Taiminen 2016: 172).

Aktualisieren und Optimieren des Lead Scoring Prozesses

Werden Lead-Scoring-Systeme nicht regelmäßig aktualisiert, kann es vorkommen, dass Leads nicht mehr korrekt bewertet werden. Aus diesem Grund ist es notwendig, das System regelmäßig zu überprüfen und dabei neue Daten und Verbesserungsvorschläge der Vertriebsmitarbeiter zu berücksichtigen (vgl. Wu et al. 2023: 16). Adobe (vgl. 2019: 32–33) empfiehlt in diesem Zusammenhang, den Prüfprozess mindestens alle drei Monate durchzuführen und dabei folgende Maßnahmen zu integrieren:

- Analysieren der Scores aller Leads und deren statistischer Verteilung
- Prüfen von Ausreißern und disqualifizierten Leads und darauf basierte Anpassung der Parameter im System
- Überprüfen des Verhaltens aktueller SALs und darauf basierte Anpassung der Parameter im System
- Einbeziehen neuer Materialien wie neu erstellter Landing Pages in das System

Bewertung des Erfolgs des Lead-Scoring-Systems

Nach der Implementierung eines Lead-Scoring-Systems ist es wichtig, den Erfolg zu überprüfen. Schließlich muss festgestellt werden, ob das Lead Scoring tatsächlich zu positiven Ergebnissen führt. Hierzu können verschiedene Key Performance Indicators (KPIs) herangezogen werden. Wu et al. (vgl. 2023: 8) haben im Rahmen ihrer Forschungsarbeit die gängigsten KPIs identifiziert, die in der Praxis des Lead Scoring verwendet werden. Diese sind in Tabelle 5 dargestellt, wobei die Spalte Count angibt, wie oft der jeweilige KPI in der untersuchten Literatur verwendet wurde.

KPI	Count
Lead conversion rate	12
Cost reduction / monetary savings	10
Number of MQLs	9
Hit rate on number of customers who buy	8
Annual revenue	7
Profit / financial gains	7
Density of profitable customers in the list	2
Response percentage	2
Customer value matrix	1
Customer overall satisfaction	1
Average time needed to qualify a lead	1
Activity level (e.g., website visits, log-ins)	1
Equilibrium percentage	1
Gain curve/score	1

Tabelle 5: KPIs im Lead Scoring
Quelle: In Anlehnung an Wu et al. 2023: 8

Daraus lassen sich die gängigsten KPIs des Lead Scorings ableiten:

- Conversion Rate
- Kosteneinsparung

- Anzahl an MQLs
- Umsatz
- Kunden-Hit-Rate
- Profit

Weitere ausgewählte KPIs, welche Adobe (vgl. 2019: 46–47) empfiehlt, sind folgende:

- Vertriebsproduktivität in Form des Umsatzes pro Vertriebsmitarbeiter
- Anteil der vom Vertriebsteam akzeptierten MQLs
- Erfolgsquote der vom Vertrieb akzeptierten Leads
- Dauer des Vertriebszyklus
- Umsatz pro Abschluss

4 Erweiterte Ansätze und Anwendungsfälle des Lead Scorings

In diesem Kapitel wird zunächst auf das prädiktive Lead Scoring eingegangen, das einen alternativen Ansatz zum traditionellen Lead Scoring darstellt. Darüber hinaus werden das produktbasierte Lead Scoring und das Account Based Scoring als erweiterte Anwendungsfälle vorgestellt. Abschließend werden noch einige spezielle Anwendungsfälle genannt, die in der analysierten Literatur vorgefunden wurden.

4.1 Prädiktives Lead Scoring

Ein alternativer Ansatz zum konventionellen Lead Scoring ist das prädiktive Lead Scoring, das dem Anwendungsbereich Predictive Analytics zugeordnet wird. Predictive Analytics fasst verschiedene mathematische und statistische Techniken zusammen, die dazu dienen, Muster in Daten zu erkennen und auf dieser Grundlage Prognosen für zukünftige Ereignisse zu treffen. Das prädiktive Lead Scoring setzt dabei auf maschinelles Lernen und statistische Modelle, um die Wahrscheinlichkeit zu prognostizieren, mit der ein Lead in einen tatsächlichen Kunden umgewandelt wird (vgl. Swani und Tyagi 2017; Nygard und Mezei 2020: 1441). Aktuelle Entwicklungen deuten darauf hin, dass prädiktive Lead-Scoring-Modelle gegenüber traditionellen Modellen zunehmend bevorzugt werden. Die vorliegende Literatur legt zudem nahe, dass die Integration von maschinellem Lernen in den Lead-Scoring-Prozess zu verbesserten Ergebnissen führt (vgl. Wu et al. 2023: 1; Duncan und Eklan 2015: 1751–1758).

Unterschiede zwischen dem traditionellen und dem prädiktiven Lead Scoring

In diesem Abschnitt werden die Unterschiede zwischen dem traditionellen und dem prädiktiven Lead-Scoring-Ansatz sowie die Ineffizienzen im traditionellen Lead Scoring näher

erläutert, die durch die Implementierung des prädiktiven Lead Scoring behoben werden können.

	Traditional Lead Scoring	Predictive Lead Scoring
Rules	Subjective rules established by expert marketers	Detected by Machine-Learning-Algorithms
Supervision	Requires manual supervision and regular adjustments and updates	Minimal supervision
Data size	Small datasets and limited processing power	Large datasets (accuracy increase with training data size)
Result	Lead scores	Conversion probability

Tabelle 6: Unterschiede im traditionellen und prädiktiven Lead Scoring
Quelle: In Anlehnung an Jadli et al. 2022: 435

Vorhersagen statt Lead Scores

Wie der Tabelle 6 zu entnehmen ist, besteht ein wesentlicher Unterschied zwischen dem prädiktiven und dem traditionellen Lead Scoring darin, dass am Ende des Prozesses kein Score steht. Welches Ergebnis am Ende des Lead-Scoring-Prozesses steht, hängt dabei von der Wahl des Algorithmus des maschinellen Lernens ab (vgl. Wu et al. 2023: 9–13). In der untersuchten Literatur wurden folgende mögliche Vorhersagen ermittelt:

- Vorhersage der Lead-Konvertierung: Durch den Einsatz von Klassifikationsalgorithmen können Leads in zwei Kategorien eingeteilt werden – solche, die voraussichtlich konvertieren werden, und solche, die nicht konvertieren werden (vgl. Bohanec et al. 2015: 338–352).
- Priorisierung von Leads: Der Einsatz von maschinellem Lernen ermöglicht es, Leads anhand ihrer Konvertierungswahrscheinlichkeit zu klassifizieren und sie anschließend entsprechend ihrer Priorität zu bearbeiten (vgl. Duncan und Eklan 2015: 1751–1758; D'Haen et al. 2016: 69–78).
- Konversionswahrscheinlichkeit: Durch den Einsatz von Algorithmen der Kategorie "Regression" kann die Wahrscheinlichkeit vorhergesagt werden, mit der einzelne Kontakte zu Kunden werden (vgl. Espadinha-Cruz et al. 2021: 1–14).
- Feature-Importance-Ausgabe: Einige Algorithmen ermöglichen auch die Darstellung der Feature Importance, die aufzeigt, welche Merkmale oder Variablen den größten Einfluss auf die Vorhersagen haben. Dies kann nützlich sein, um zu verstehen, welche Faktoren die Konversionswahrscheinlichkeit am stärksten beeinflussen (vgl. Bohanec et al. 2017: 416–428). Somit besteht die Möglichkeit, diese Informationen zu nutzen, um das Punktesystem in einem traditionellen Lead-Scoring-System zu

entwickeln oder das Vorgehen innerhalb eines prädiktiven Systems besser nachvollziehbar zu machen.

Geringerer Aktualisierungsaufwand

Wie in Kapitel 3.3 festgestellt, muss regelmäßig Zeit in die Wartung traditioneller Lead-Scoring-Systeme investiert werden. Im Gegensatz dazu benötigen prädiktive Systeme einen geringeren Wartungsaufwand (siehe Tabelle 6).

Datengestützte Entscheidungsgrundlage statt subjektiver Experteneinschätzungen

Wie bereits erörtert, variieren die entscheidenden Parameter im Lead Scoring je nach Unternehmen. Im herkömmlichen Ansatz stützt man sich auf Expertenannahmen, um die einzelnen Parameter und ihre Gewichtung festzulegen. Diese Vorgehensweise führt dazu, dass das Lead Scoring nicht auf tatsächlichen Daten basiert, was wiederum zu ungenauen Ergebnissen führen kann. Im Kontrast dazu basiert das prädiktive Lead Scoring vollständig auf empirischen Daten, wodurch keine subjektiven Einschätzungen in das Modell einfließen (vgl. Duncan und Eklan 2015: 1751–1758). Dadurch lässt sich die Herausforderung der Identifizierung der Kaufwahrscheinlichkeitsfaktoren effizienter bewältigen.

Darüber hinaus können prädiktive Lead-Scoring-Modelle Zusammenhänge erkennen, die zu komplex sind, um von Experten erkannt zu werden (vgl. Wu et al. 2023: 15). Außerdem können auch nicht-lineare Effekte berücksichtigt werden, die im traditionellen Lead Scoring nicht durch die Scorecard abgebildet werden können (vgl. Duncan und Eklan 2015: 1752).

Prädiktive Lead-Scoring-Modelle zeichnen sich im Vergleich zum traditionellen Ansatz obendrein durch eine geringere Abhängigkeit von der Menge an verhaltensbasierten Daten aus. Im traditionellen Lead Scoring werden potenzielle Leads oft erst dann identifiziert, wenn ausreichend Aktionen durchgeführt wurden, um einen hohen Score zu erreichen. Bei prädiktiven Modellen kann die Abhängigkeit von der Menge verhaltensbasierter Daten jedoch reduziert werden, wodurch auch bei von vornherein kaufbereiten Kontakten eine genauere Einschätzung der Kaufwahrscheinlichkeit möglich ist (vgl. Duncan und Eklan 2015: 1752).

Eignung zur Auswertung großer Datenmengen

Prädiktive Modelle bieten den Vorteil, besonders effektiv bei der Analyse und Auswertung großer Datenmengen zu sein (siehe Tabelle 6). Im Vergleich dazu stoßen Experten bei der Entwicklung von Scorecards für traditionelle Modelle an ihre Grenzen, da die Berücksichtigung zahlreicher Parameter und deren Gewichtung bei großen Datenmengen zu komplex werden kann.

Geringere Transparenz bei prädiktiven Modellen

Ein Nachteil prädiktiver Lead-Scoring-Algorithmen liegt darin, dass die zugrundeliegenden Modelle als "Black Box" fungieren und somit nicht nachvollziehbar ist, was innerhalb eines

solchen Modells geschieht (vgl. Wu et al. 2023: 17). Diese Intransparenz kann sich negativ auf die Zusammenarbeit zwischen Marketing und Vertrieb auswirken, da die Modelle keinen Raum für Diskussionen bieten. Wenn dem Vertrieb das Vertrauen in die Qualität der auf Basis des Modells übermittelten Leads fehlt, kann die Effizienz des Prozesses erheblich beeinträchtigt werden. Es gibt jedoch Methoden, um das Vorgehen innerhalb der Algorithmen des maschinellen Lernens besser nachvollziehbar zu machen und so das Problem der Intransparenz zu reduzieren (vgl. Bohanec et al. 2017: 416–428).

Entwickeln eines prädiktiven Lead-Scoring-Modells

In diesem Abschnitt werden die einzelnen Schritte zur Erstellung eines prädiktiven Lead-Scoring-Modells untersucht. Prädiktives Lead Scoring ist ein konkreter Anwendungsfall aus dem Bereich Predictive Analytics und Machine Learning und fällt somit in den Bereich des Data Minings (vgl. Elkan 2013: 7; Swani und Tyagi 2017: 5–10; Jo 2021: 19). Aus diesem Grund wird in dieser Arbeit bei der Entwicklung eines prädiktiven Lead-Scoring-Systems nach dem Cross-Industry Standard Process for Data Mining (CRISP-DM) vorgegangen. Der CRISP-DM ist eine generische und bereichsübergreifende Methodik für Data Mining, die sowohl Einsteigern als auch Experten einen Leitfaden für die Durchführung eines Data Mining Projektes bietet. Dieser wurde bereits Mitte 1996 entwickelt (vgl. Shearer 2000: 14–19), wird jedoch nach wie vor noch als Standard im Bereich Data Mining angesehen (vgl. Martinez-Plumed et al. 2021: 3048; Abbasi et al. 2016: 13). Der Prozess wird hierbei in sechs Phasen gegliedert: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation und Deployment (vgl. Shearer 2000: 14–19).

<u>Business Understanding</u>

In dieser Phase werden Projektziele definiert, Geschäftsziele verstanden und ein Plan entwickelt. Schlüsselschritte umfassen die Bestimmung von Geschäftszielen, die Bewertung der Situation, die Festlegung von Datenanalysezielen und die Erstellung eines Projektplans (vgl. Shearer 2000: 14–19).

<u>Data Understanding</u>

Diese Phase beginnt mit der initialen Datensammlung und umfasst Schritte wie die Beschreibung der Daten, die Exploration durch Abfragen und Visualisierungen sowie die Überprüfung der Datenqualität (vgl. Shearer 2000: 14–19):

- Sammeln der verfügbaren Daten: In diesem Schritt werden die notwendigen Daten erworben.
- Data Description: Die Eigenschaften der zuvor erworbenen Daten werden untersucht, wobei Informationen wie das Datenformat, die Menge, die Anzahl der Datensätze und Felder in jeder Tabelle gewonnen werden.

- Data Exploration: Durch Abfragen, Visualisierung und Berichterstellung werden detaillierte Einblicke in die Daten gewonnen. Erste Erkenntnisse und Hypothesen werden darauf basierend festgehalten.
- Prüfen der Datenqualität: Die Untersuchung der Datenqualität erfolgt, um sicherzustellen, dass die Daten vollständig sind und keine fehlenden Werte enthalten. Zusätzlich wird die Plausibilität der Daten überprüft.

Data Preparation

Diese Phase umfasst alle Aktivitäten zum Aufbau des endgültigen Datensatzes oder der Daten, die später in das Modell einfließen. Diese werden im Folgenden beschrieben:

- Datenauswahl: Die Entscheidung, welche Daten für die Analyse verwendet werden, basiert auf Kriterien wie der Relevanz für die Ziele der Datenanalyse, der Qualität und technischen Einschränkungen (vgl. Shearer 2000: 16).
- Datenbereinigung: Die Datenbereinigung ist ein wesentlicher Schritt im Data Mining, da die Qualität der Ergebnisse von der Sauberkeit der Daten abhängt (vgl. Shearer 2000: 16). Im Kontext des maschinellen Lernens gibt es Modelle, die mit fehlenden Daten umgehen können, während andere auf vollständige Daten angewiesen sind. Daher ist es notwendig, fehlende Werte in den Daten zu behandeln, indem die entsprechenden Zeilen entweder gelöscht oder fehlende Werte durch geeignete Techniken wie die Ergänzung durch den Mittelwert oder den Median einer Spalte ersetzt werden (vgl. Kumar und Reinartz 2018: 139; Elkan 2013: 19–20; Shearer 2000: 16). Neben der Behandlung fehlender Werte umfasst die Datenbereinigung auch die Identifikation und Eliminierung von Ausreißern sowie die Korrektur fehlerhafter Werte (vgl. Kumar und Reinartz 2018: 139–140; Cleff 2019: 24; Kuhn und Johnson 2013: 33–35). Im Bereich des Lead Scorings könnte beispielsweise darauf geachtet werden, Leads mit ungewöhnlich hoher Aktivität auszusortieren, da diese möglicherweise von Systemtestern stammen und die Analyseergebnisse verfälschen könnten. Dies trägt dazu bei, die Qualität des Lead Scorings zu verbessern, da Extremwerte die Ergebnisse nicht beeinflussen. Ein weiterer Aspekt der Datenbereinigung beim maschinellen Lernen ist die Datenreduktion. Hierbei wird versucht, die Datenmenge, die in das Modell einfließt, zu reduzieren. Dies kann durch das Entfernen von Spalten geschehen, die keine relevanten Informationen für das Modell liefern. Ein Beispiel hierfür ist die Eliminierung redundanter Variablen, die stark mit anderen Variablen korrelieren und somit die Analyseergebnisse negativ beeinflussen könnten (vgl. D'Haen und van den Poel 2013: 548; Kuhn und Johnson 2013: 35).

- Datenkonstruktion: Nach der Datenbereinigung werden im Rahmen der Datenkonstruktion gegebenenfalls neue Datensätze oder aus bestehenden Datensätzen abgeleitete Attribute entwickelt. Zweck dieser abgeleiteten Attribute ist es, den Modellierungsprozess zu erleichtern beziehungsweise den Modellierungsalgorithmus zu unterstützen (vgl. Shearer 2000: 16). Grundsätzlich sind textbasierte Variablen im Bereich des maschinellen Lernens nicht geeignet, um Vorhersagen zu treffen. Daher werden textbasierte Daten im Prozess der Datenkonstruktion kodiert, indem sie in numerische Werte umgewandelt werden. Hierbei werden verschiedene Kodierungstechniken angewendet (vgl. Kuhn und Johnson 2013: 47–48; D'Haen und van den Poel 2013: 548; Uhlemann 2015: 10). Ein Beispiel im Kontext des Lead Scorings wäre die Kodierung der Spalte "Aktion" in einer Liste, die alle Aktionen aller Leads enthält. Angenommen, die Spalte "Aktion" enthält die vier eindeutigen Aktionen "Besuch Preisseite A", "Besuch Preisseite B", "Download PDF A" und "Download PDF B", dann wäre eine mögliche Vorgehensweise, für jeden Kunden zwei Spalten anzulegen: Eine mit der Anzahl der Besuche von Preisseiten und eine mit der Anzahl der Downloads von PDFs.
- Datenintegration: Bei der Datenintegration werden Informationen aus mehreren Tabellen oder Datensätzen zusammengeführt, um eine Tabelle zu erstellen, die alle für das Modell erforderlichen Informationen enthält. Zur Datenintegration gehören auch Aggregationen. Aggregationen beziehen sich auf Operationen, bei denen neue Werte berechnet werden, indem Informationen aus mehreren Datensätzen und/oder Tabellen zusammengefasst werden. Ein Beispiel für eine Aggregation könnte die Transformation einer Tabelle von Kundenkäufen sein, in der es einen Datensatz für jeden getätigten Kauf gibt. Diese Tabelle könnte in eine neue Form umgewandelt werden, in der es für jeden Kunden einen Datensatz gibt. Die Felder dieser neuen Tabelle könnten Informationen wie die Anzahl der getätigten Käufe, den durchschnittlichen Kaufbetrag, den Prozentsatz der mit Kreditkarte bezahlten Bestellungen und andere aggregierte Werte enthalten (vgl. Shearer 2000: 16–17).
- Datenformatierung: Im Zusammenhang mit Data Mining kann es notwendig sein, das Format oder die Struktur von Daten anzupassen. Diese Anpassungen können einfach sein, wie das Entfernen unzulässiger Zeichen aus Zeichenketten oder das Kürzen auf eine maximale Länge, oder komplexer, wie die Umstrukturierung von Informationen. Solche Änderungen sind mitunter erforderlich, um die Daten für die Anwendung eines bestimmten Modellierungswerkzeugs geeignet zu machen (vgl. Shearer 2000: 17). Bei der Entwicklung von Modellen des maschinellen Lernens ist es im Rahmen der Datenformatierung wichtig, die Daten in Test- und Trainingsdaten zu unterteilen, um

später eine genaue Beurteilung der Leistungsfähigkeit des Modells zu ermöglichen. Hierbei werden Modelle zunächst auf der Grundlage von Trainingsdaten trainiert. Die Testdaten werden dann verwendet, um die Leistung der trainierten Modelle zu bewerten, indem sie anhand unbekannter Daten validiert werden. Diese Aufteilung ermöglicht es, das sogenannte Overfitting zu erkennen und sicherzustellen, dass ein Modell nicht nur die Trainingsdaten, sondern auch bisher unbekannte Daten korrekt vorhersagen kann. Dies erhöht die Zuverlässigkeit des Modells in realen Situationen (vgl. Kuhn und Johnson 2013: 60–61). Darüber hinaus kann es sinnvoll sein, die Daten zu skalieren und zu zentrieren. Bei der Zentrierung kann zum Beispiel von jeder Variablen der Mittelwert abgezogen werden, so dass der neue Mittelwert null ist. Beim Skalieren können die einzelnen Variablen durch die Standardabweichung innerhalb der Spalte dividiert werden. Dieses Vorgehen kann die Vorhersage von Modellen verbessern, da die einzelnen Spalten nun auf einer gemeinsamen Skala liegen (vgl. Cleff 2019: 20; Kuhn und Johnson 2013: 30–31).

Modeling

In dieser Phase werden verschiedene Modelle ausgewählt und angewendet, wobei die Parameter jedes Modells auf optimale Werte kalibriert werden. Dazu werden die folgenden Schritte durchgeführt:

- Modellauswahl: In diesem Arbeitsschritt werden ein oder mehrere Modelle ausgewählt, die im Rahmen des Projektes verwendet werden sollen. Prädiktive Lead-Scoring-Algorithmen können dabei aus den Kategorien Klassifikation, Regression und Clustering stammen, wobei Klassifikations-Algorithmen am häufigsten verwendet werden. Regressionsmodelle liefern numerische Werte auf Basis von Parametern, die in das Modell eingegeben werden. Dies kann zum Beispiel die Wahrscheinlichkeit sein, mit der ein Lead konvertiert. Klassifikationsmodelle hingegen ordnen die in das Modell eingegebenen Objekte vordefinierten Kategorien zu, beispielsweise der Gruppe "Lead wird konvertieren" und der Gruppe "Lead wird nicht konvertieren". Clustering-Modelle funktionieren ähnlich wie Klassifikationsmodelle, allerdings arbeiten sie ohne vordefinierte Kategorien, sondern bilden die einzelnen Gruppen selbst (vgl. Wu et al. 2023: 9–13). In der untersuchten Literatur wurden Algorithmen des maschinellen Lernens wie Linear Regression, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine und neuronale Netze verwendet. Die besten Ergebnisse wurden dabei unter Verwendung des Random-Forest-Algorithmus erzielt, der ein Modell der Kategorie Klassifikation ist (vgl. Bohanec et al. 2017: 416–428; Gokhale und Joshi 2018: 279–291; Jadli et al. 2022: 433–443; Nygard und Mezei 2020: 1439–1448).

- Modellentwicklung: Nach der Modellauswahl folgt der Schritt der Modellentwicklung. Beim maschinellen Lernen wird in diesem Schritt unter anderem das Hyperparameter-Tuning durchgeführt. Beim Hyperparameter-Tuning werden für jedes ausgewählte Modell systematisch verschiedene Einstellungen ausprobiert, die die Modellleistung beeinflussen. Dabei werden die Hyperparameter ermittelt, welche die beste Leistung erzielen. Das Hauptziel dieses Prozesses ist es, die Hyperparameter so anzupassen, dass das Modell gute Ergebnisse auf bisher unbekannten Daten liefert (vgl. Kuhn und Johnson 2013: 63, 73).
- Modellbewertung: Die Evaluierung von Modellen des maschinellen Lernens erfordert eine präzise Analyse ihrer Leistung, wobei die Konfusionsmatrix als Instrument dient. Diese Matrix bietet einen detaillierten Überblick über die Vorhersagen eines Modells im Vergleich zu den tatsächlichen Ergebnissen. Die Konfusionsmatrix enthält vier Kategorien: True Positives (TP) für korrekt vorhergesagte positive Ergebnisse und True Negatives (TN) für korrekt vorhergesagte negative Ergebnisse. False Positives (FP) und False Negatives (FN) beschreiben die Anzahl der falsch vorhergesagten positiven oder negativen Ergebnisse. Ein Beispiel für eine Konfusionsmatrix findet sich in Tabelle 7. Eine Metrik zur Bewertung der Leistung eines maschinellen Lernmodells, die aus der Konfusionsmatrix abgeleitet wird, ist die Genauigkeit (Accuracy). Diese gibt den Anteil der korrekten Vorhersagen an, indem die Anzahl der TP- und FP-Vorhersagen durch die Gesamtanzahl der Vorhersagen dividiert wird. Die Präzision gibt den Anteil der korrekten Vorhersagen im Verhältnis zu allen positiven Vorhersagen an. Die Sensitivität gibt den Anteil der korrekten positiven Vorhersagen im Verhältnis zur Anzahl der tatsächlich positiven Werte an. Die 1-Spezifität gibt den Anteil an falsch positiven Vorhersagen im Verhältnis zu allen falschen Vorhersagen an (vgl. Nygard und Mezei 2020: 1444; Von der Hude 2020: 149–152).

		Predicted	
		Positive	Negative
Truth	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Tabelle 7: Konfusionsmatrix
Quelle: In Anlehnung an Elkan 2013: 49

Evaluation

Vor der endgültigen Bereitstellung des Modells ist es wichtig, das Modell zu evaluieren und den Konstruktionsprozess zu überprüfen. Die wichtigsten Schritte sind die Bewertung der Ergebnisse im Hinblick auf die Erreichung der Geschäftsziele, die Überprüfung des Prozesses auf Schwachstellen und die Festlegung der nächsten Schritte (vgl. Shearer 2000: 17–18).

Deployment

Mit der Erstellung und Evaluierung des Modells ist das Projekt in der Regel nicht abgeschlossen. Das erworbene Wissen muss organisiert und aufbereitet werden, damit es genutzt werden kann. Die wichtigsten Schritte in diesem Zusammenhang sind die Planung des Einsatzes, die Überwachung und Pflege und die Erstellung des Abschlussberichtes (vgl. Shearer 2000: 18).

4.2 Produktbasiertes Scoring

Das produktbasierte Scoring ist eine Form der Leadbewertung, die es Unternehmen ermöglicht, das Interesse potenzieller Kunden an verschiedenen Produkten zu messen. Dieser Ansatz geht über das allgemeine verhaltensbasierte Scoring hinaus, das sich auf das Interesse am Unternehmen als Ganzes konzentriert. Stattdessen werden beim produktbasierten Scoring verschiedene Scores für unterschiedliche Produkte erstellt, um detailliertere Informationen zu sammeln. Wenn ein Lead den Schwellenwert für einen Produkt-Score überschreitet, kann dieser an das für dieses Produkt zuständige Vertriebsteam weitergeleitet werden. Beim produktbasierten Scoring können Produkt-Scores nicht nur für einzelne Produkte, sondern auch für übergeordnete Produktgruppen gebildet werden (vgl. Adobe 2019: 29). Daher kann das Scoring theoretisch auch auf Basis der einzelnen Geschäftsbereiche erfolgen. Welche Gliederung sinnvoll ist, kann im Vorfeld mit den einzelnen Vertriebsteams abgestimmt werden. Produktbasiertes Scoring kann beispielsweise in einem Unternehmen eingesetzt werden, das ERP-Software, CRM-Software und eine Supply-Chain-Management-Software vertreibt. Wird in diesem Fall nur ein Score verwendet, ist es nicht möglich, das Interesse der Leads ausschließlich in Bezug auf das Produkt CRM-Software zu messen und anschließend die MQLs für CRM-Software an das dafür vorgesehene Vertriebsteam zu übergeben. Die Möglichkeit des produktbasierten Scorings wird von mehreren Anbietern von MAS und CRM-Systemen angeboten (vgl. Oracle o. D.; Salesforce o. D.; InvestGlass 2023).

In der Praxis empfiehlt es sich, mit dem Lead Scoring klein und mit reduzierter Komplexität zu beginnen und das System sukzessive auszubauen (vgl. Schoepf 2021: 283; Auerochs 2021). Auch hier kann das Produkt Scoring helfen, da es einzelnen Produktbereichen ermöglicht, isolierte Scoring-Systeme zu entwickeln, anstatt ein komplexes

unternehmensweites System aufbauen zu müssen. So kann das Lead-Scoring-System bei Bedarf zunächst in einer einzelnen Abteilung getestet werden, bevor es unternehmensweit eingesetzt wird.

4.3 Account Based Scoring

Neben dem produktbasierten Scoring ist das Account Based Scoring (ABS) ein weiterer fortgeschrittener Anwendungsfall des Lead Scorings. ABS ist Teil des Account Based Marketings und bewertet anstelle einzelner Leads die den Leads übergeordneten Unternehmen. Ziel ist somit nicht die Qualifizierung einzelner Leads zum MQL, sondern die Qualifizierung eines Unternehmens beziehungsweise eines Accounts zum Marketing Qualified Account (MQA) (vgl. Day und Wei Shi 2020: 18–19).

Das Scoring auf Unternehmensebene kann effektiver sein, da bei größeren Transaktionen im Business-to-Business (B2B)-Bereich in der Regel mehrere Personen am Kaufprozess beteiligt sind. Daher bietet die Betrachtung eines einzelnen Leads keinen ausreichenden Einblick mehr, um auf das Kaufinteresse des Unternehmens zu schließen (vgl. Day und Wei Shi 2020: 16; Schuster 2022: 203). So kann es beispielsweise sein, dass von den einzelnen Leads des Unternehmens keiner den festgelegten Score überschreitet, die Leads zusammen jedoch einen sehr hohen Score aufweisen, was ein klares Indiz für das Kaufinteresse des Unternehmens ist (vgl. Adobe 2019: 30).

Zur Entwicklung eines ABS-Modells werden zunächst qualitativ hochwertige Unternehmen untersucht, die in der Vergangenheit vom Vertrieb akzeptiert wurden. Auf dieser Basis können dann Kriterien abgeleitet werden, die auf eine hohe Account-Qualität schließen lassen. Anschließend werden die Scorecards für das ABS erstellt. Explizite Parameter können hier beispielsweise folgende sein (vgl. Day und Wei Shi 2020: 20):

- Anzahl der Mitarbeiter
- Umsatz
- Standort
- Branche
- Technologieprofil
- Einstellungsinformationen
- Produktinformationen
- Finanzierung
- Web-Ranking
- Präsenz in sozialen Medien

Der implizite Score ergibt sich aus der Summe der Punkte für die Aktivitäten, die alle dem Unternehmen zugeordneten Leads gesammelt haben. Werden die festgelegten Schwellenwerte überschritten, erhält das Unternehmen den Status MQA und wird an den Vertrieb übergeben (vgl. Day und Wei Shi 2020: 19–20).

In Zuge des ABS kann auch ein prädiktiver Lead-Scoring-Ansatz angewandt werden (vgl. Heinzelbecker 2021a: 394). Allerdings ist zu beachten, dass beim ABS eine geringere Anzahl an Accounts bewertet wird (siehe Tabelle 8). Da für das Entwickeln eines prädiktiven Lead-Scoring-Systems ausreichend Daten zur Verfügung stehen müssen (vgl. Elkan 2013: 8), muss folglich sichergestellt werden, dass ausreichend Daten aus der Vergangenheit zur Verfügung stehen, um den Algorithmus zu trainieren.

	B2C	B2B
Target audience	End User	Enterprise
Target market size	Large	Smaller
Sales volume	Low	High
Decision making	By the consumer	By a committee
Risk	Low	High
Purchasing process	Short	Longer
Consumer decision	Emotional	Rational
Usage of mass media for promotion	Common	Often avoided

Tabelle 8: Unterschiede im B2C- und B2B-Markt
Quelle: In Anlehnung an Saha et al. 2014: 295–297

4.4 Lead Scoring ohne Vertriebsteam

Innerhalb der wissenschaftlichen Literatur wird nahezu ausschließlich die Übergabe an ein Vertriebsteam als Aktion beim Überschreiten eines Scores-Schwellenwertes genannt. Lediglich Zumstein et al. (vgl. 2023: 35) erwähnen die Durchführung spezieller Kommunikationsmaßnahmen auf Basis des Lead Scores und deuten damit an, dass MQLs nicht zwingend an den Vertrieb übergeben werden müssen. Da in der Praxis auch Unternehmen ohne Vertriebsmannschaft existieren, beispielsweise im Bereich E-Commerce, wurde in Suchmaschinen nach weiteren Möglichkeiten recherchiert, um Maßnahmen auf Basis des Scores im Lead Scoring durchzuführen. Dabei wurden folgende identifiziert:

- Lead-Segmentierung: Eine Möglichkeit besteht darin, Leads auf Basis ihrer Scores dynamisch Segmenten zuzuordnen, beispielsweise "Kalte Leads", "Lauwarme Leads", "Warme Leads" und "Heiße Leads". Dies ermöglicht nicht nur eine bessere

Strukturierung der Leads, sondern ermöglicht auch die Durchführung angepasster Marketingmaßnahmen zu einem späteren Zeitpunkt (vgl. Brevo 2023).

- Versenden von personalisierten Lead-Nurturing-Inhalten, Angeboten und Calls To Action (CTAs): In Unternehmen ohne eigene Vertriebsmannschaft können spezielle Angebote und CTAs an MQLs versendet werden, um diese in Kunden umzuwandeln, nachdem ihr Score den Schwellenwert überschritten hat. Zudem kann der Lead Score im Rahmen des Lead Nurturings dazu verwendet werden, sicherzustellen, dass Leads die passenden Inhalte zur richtigen Zeit erhalten. Dies ermöglicht beispielsweise, warme Leads mit speziellen Inhalten zu versorgen, um sie in heiße Leads zu verwandeln (vgl. Hufford 2021; Ghorbel 2023). Im produktbasierten Lead Scoring können Inhalte, Angebote und CTAs zudem auf die Produktgruppen zugeschnitten werden, für die sich ein Lead interessiert (vgl. faraday.ai o. D.).
- Targeting in sozialen Median: Leads, die einen vordefinierten Score überschreiten, können zusätzlich über soziale Medien angesprochen werden, um die Wahrscheinlichkeit eines Kaufs zu steigern (vgl. Brevo 2023).

4.5 Weitere Anwendungsfälle

Neben den in diesem Kapitel vorgestellten Methoden, die in mehreren wissenschaftlichen Publikationen untersucht wurden, gibt es auch spezifischere Anwendungsfälle, die in vereinzelten Publikationen untersucht wurden. So wurde beispielsweise das prädiktive Lead Scoring zur Identifikation von abwanderungsgefährdeten Bestandskunden vorgeschlagen und eingesetzt (vgl. Buckinx und van den Poel 2005: 264; Simmoleit 2023). Darüber hinaus konstruierten Kim und Street (vgl. 2004: 215–228) ein Modell, das für Leads den Gewinn berechnete, der durch den Versand von Direktwerbung erzielt werden kann. Somit wurde Direktwerbung nur an die Leads gesendet, für die ein hoher Profit vorhergesagt wurde, um dadurch den Gewinn zu maximieren.

5 Alternative und ergänzende Methoden zum Lead Scoring

5.1 RFM-Analyse

Die RFM-Analyse ist eine Methode aus dem Bereich CRM. RFM steht dabei für "Recency", "Frequency" und "Monetary". Die Methode dient der Segmentierung von Kunden anhand ihrer letzten Transaktionen (Recency), der Häufigkeit ihrer Transaktionen (Frequency) und des monetären Wertes ihrer Einkäufe (Monetary). Ziel ist es, die Kunden nach ihrer Qualität in Gruppen einzuteilen, um sie mit gezielteren Marketingstrategien und -aktivitäten anzusprechen (vgl. Kumar und Reinartz 2018: 103–111). Dazu werden zunächst in einer

Tabelle (sehe Tabelle 9), in der alle Einkäufe der Kontakte aufgelistet sind, für jeden Eintrag Punkte vergeben. Die Punktevergabe wird vorher festgelegt. Im Beispiel werden 100 Punkte für Käufe vergeben, die in den letzten zwei Monaten getätigt wurden, 50 Punkte für Käufe, die in den letzten sechs Monaten getätigt wurden, 15 Punkte für Käufe, die in den letzten neun Monaten getätigt wurden, und fünf Punkte für Käufe, die in den letzten zwölf Monaten getätigt wurden. Die Frequency-Punkte werden berechnet, indem für jeden Kauf in den letzten zwölf Monaten sechs Punkte vergeben werden, wobei maximal 30 Punkte erreicht werden können. Die Monetary-Punkte betragen in diesem Beispiel zehn Prozent des Einkaufswertes, wobei maximal 75 Punkte erreicht werden können.

Customer	Recency Points	Frequency Points	Monetary Points
John	100	6	12
John	50	6	36
John	5	6	18
Smith	15	6	75
Mary	100	6	27
Mary	50	6	21
Mary	25	6	24
Mary	5	6	12

Tabelle 9: RFM-Tabelle zur Punktvergabe pro Einkauf
Quelle: In Anlehnung an Kumar und Reinartz 2018: 111

Anschließend werden die Punkte der einzelnen Käufe pro Kontakt aufsummiert. Werden nun der Recency-Score, der Frequency-Score und der Monetary-Score addiert, so erhält man den RFM-Score (siehe Tabelle 10).

Customer	Recency Score	Frequency Score	Monetary Score	RFM-Score
John	165	18	66	249
Smith	15	6	75	96
Mary	180	24	84	288

Tabelle 10: Errechnung des RFM-Scores
Quelle: In Anlehnung an Kumar und Reinartz 2018: 103–111

Die RFM-Analyse und das Lead Scoring unterscheiden sich folglich in ihren Anwendungsbereichen. Die RFM-Analyse zielt darauf ab, bestehende Kunden zu bewerten und kann somit keine Leads qualifizieren, die noch nicht gekauft haben. Das Lead Scoring hingegen kann auch neue Leads bewerten.

5.2 Produktempfehlungssysteme

Produktempfehlungssysteme sind leistungsstarke Werkzeuge, die insbesondere von E-Commerce-Unternehmen eingesetzt werden, um Nutzern relevante Produkte vorzuschlagen (vgl. Hu und Zhang 2012: 1). Die Systeme arbeiten in zwei Schritten. Zunächst werden die Aktivitäten und Interessen der Benutzer analysiert. Zweitens versucht es, eine Gruppe von Artikeln zu finden, die für die Nutzer von Interesse sein könnten (vgl. Sharma et al. 2021: 1).

Zu diesem Zweck werden hauptsächlich inhaltsbasierte Systeme, kollaborative Filtersysteme und hybride Ansätze beider Systeme verwendet. Wie in Abbildung 3 dargestellt, analysieren inhaltsbasierte Systeme die Eigenschaften und Merkmale der Produkte selbst. Sie verwenden Informationen wie Schlüsselwörter und Kategorisierungen, um Produkte zu empfehlen, die den Produkten ähneln, für die sich Nutzer interessieren. Beispielsweise könnte ein inhaltsbasiertes System Nutzern, die nach formeller Kleidung suchen, weitere formelle Modeaccessoires empfehlen (vgl. Sharma et al. 2021: 3).

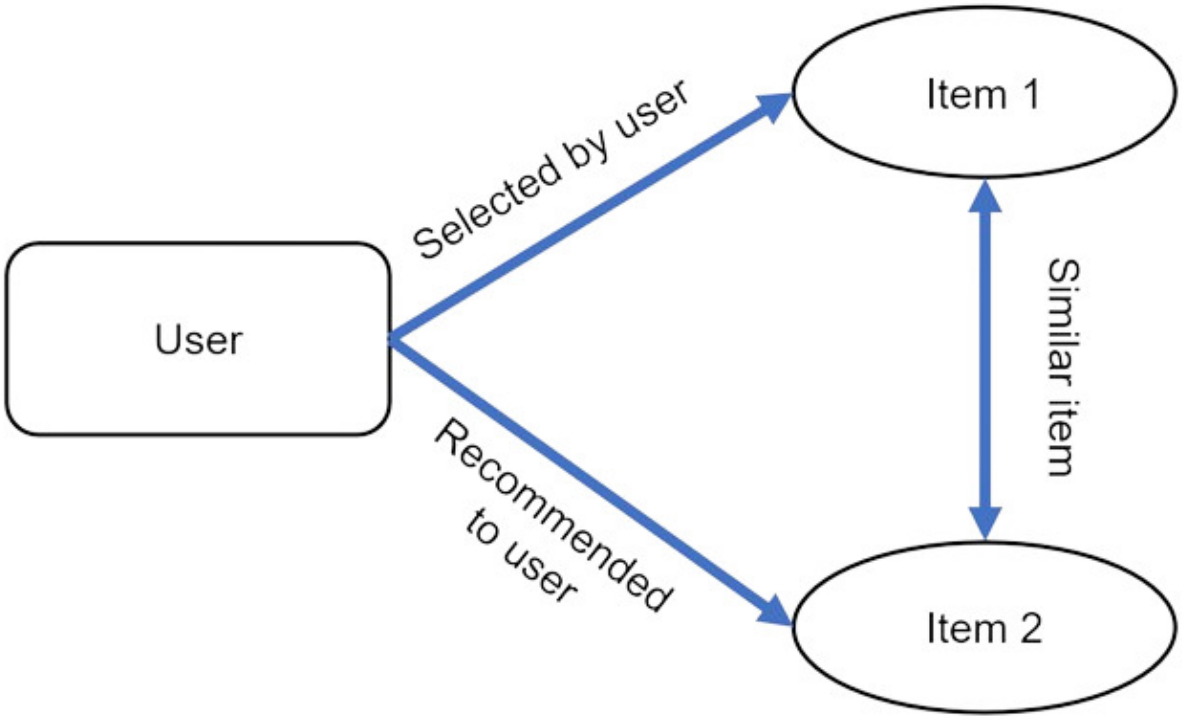

Abbildung 3: Inhaltsbasierte Produktempfehlungssysteme
Quelle: In Anlehnung an Sharma et al. 2021: 3

Bei kollaborativen Filtersystemen werden Empfehlungen auf der Grundlage von Interaktionen und Bewertungen ähnlicher Nutzer erstellt. Dazu werden die Aktivitäten der Nutzer analysiert, um ähnliche Nutzergruppen zu identifizieren. Hat Nutzer A ähnliche Präferenzen wie Nutzer B, so können Produkte, die Nutzer B gefallen haben, auch Nutzer A empfohlen werden (siehe Abbildung 4).

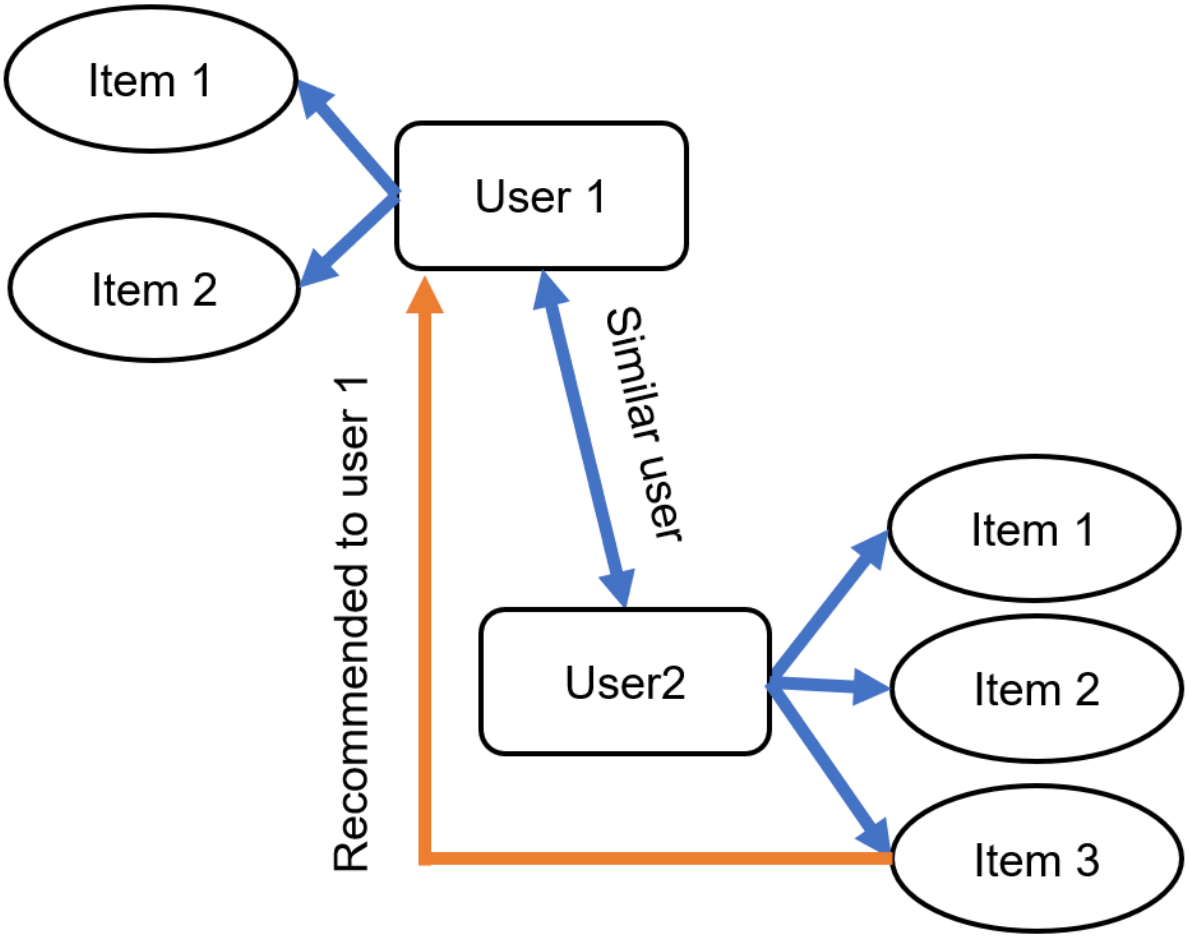

Abbildung 4: Kollaborative Filtersysteme
Quelle: In Anlehnung an Sharma et al. 2021: 4

Der hybride Ansatz ist eine Kombination aus inhaltsbasierten Systemen und kollaborativen Filtersystemen. Dadurch werden den Nutzern sowohl Artikel empfohlen, die ihren Interessen entsprechen, als auch Artikel, die von Kunden mit ähnlichem Kaufverhalten bevorzugt werden (vgl. Sharma et al. 2021: 3–4).

Algorithmen, die für die Umsetzung der zuvor genannten Systeme in Betracht gezogen werden können, umfassen beispielsweise K-Nearest-Neighbor und Matrix Factorization (vgl. Sharma et al. 2021: 4–5). Darüber hinaus sind Ansätze wie Bipartite Projection, Spanning Tree und die Anwendung der Kosinus-Ähnlichkeit relevante Methoden im Kontext von Produktempfehlungssystemen (vgl. Hu und Zhang 2012: 2).

Im Gegensatz zum Lead Scoring, das darauf abzielt, Kunden hinsichtlich ihrer Kaufwahrscheinlichkeit zu bewerten, haben Produktempfehlungssysteme somit das Ziel, den Kunden die für sie relevantesten Produkte vorzuschlagen. Trotz dieser Unterschiede können Produktempfehlungssysteme ähnliche Ziele wie das produktbasierte Lead Scoring verfolgen, da beide die Möglichkeit bieten, aus einem Sortiment verschiedener Produkte diejenigen zu identifizieren, die für einen bestimmten Lead besonders relevant sind.

6 Entwicklung eines generischen Vorgehensmodells zum Lead Scoring

Basierend auf den zuvor identifizierten Herausforderungen des Lead Scorings wird in diesem Kapitel ein generisches Vorgehensmodell zur Entwicklung eines traditionellen und eines prädiktiven Lead-Scoring-Systems vorgestellt. Diese beiden Vorgehensmodelle werden in Abbildung 5 dargestellt, wobei die Unterschiede der Modelle in Hellblau hervorgehoben werden.

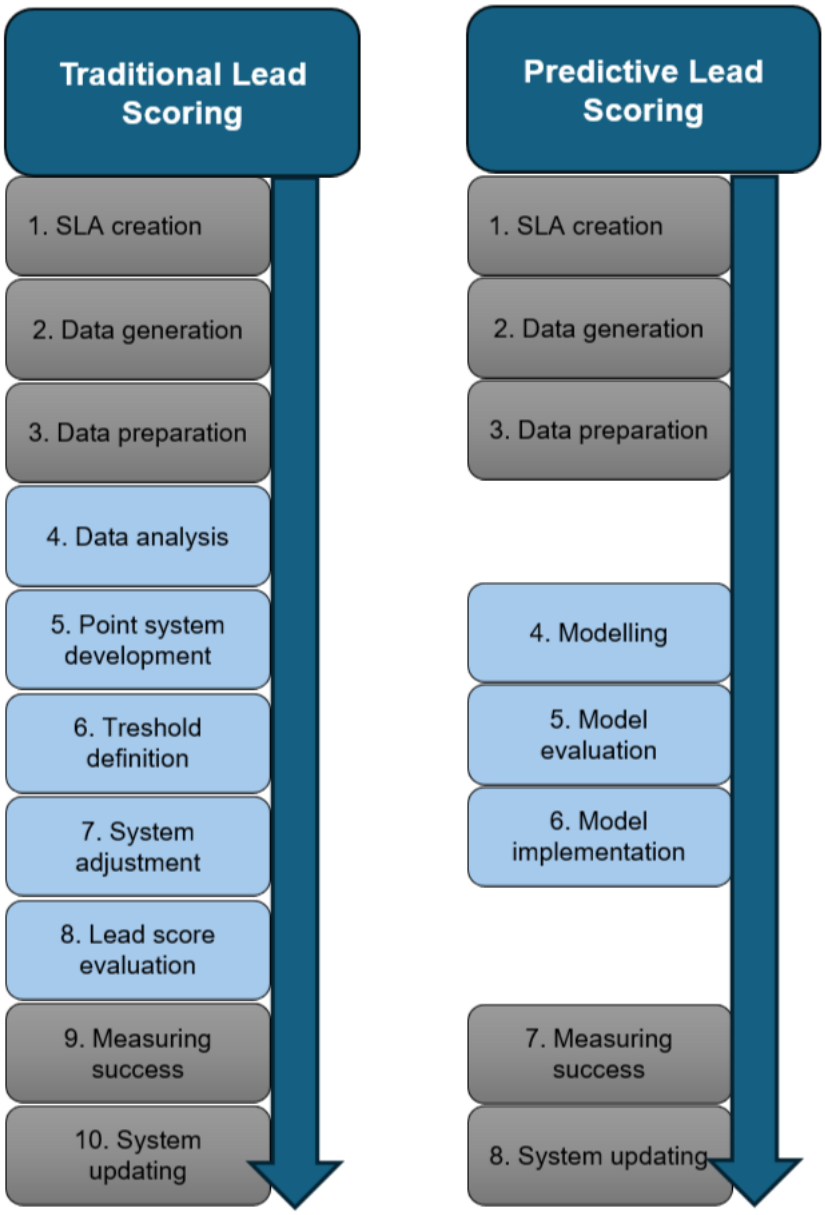

Abbildung 5: Unterschiede zwischen dem traditionellen und dem prädiktiven Vorgehensmodell zur Erstellung eines Lead-Scoring-Systems
Quelle: Eigene Darstellung

6.1 Traditionelles Lead Scoring

Erstellen eines Service Level Agreements

Wie in Kapitel 3.3 beschrieben, empfiehlt es sich, dass das Marketing und der Vertrieb im ersten Schritt die Rahmenbedingungen des Lead-Scoring-Projektes im SLA klar definieren. In diesem sollen unter anderem folgende Fragestellungen geklärt werden:

- Auf welche Zielgruppen wird das Lead Scoring angewandt?
- Wie sieht das ideale Lead-Profil aus?
- Wie werden MQLs, SALs und SQLs definiert?
- Welche Stufen durchlaufen Interessenten im Rahmen der Customer Journey?
- Welche Daten sollen gesammelt werden?
- Welche Daten müssen zwischen der MAS und anderen Systemen übertragen werden?
- Welche Methode zur Leadbewertung wird gewählt?
- Wann ist ein Lead reif, um an den Vertrieb übergeben zu werden?
- Welche Informationen zu den Leads werden bei der Übergabe an den Vertrieb übermittelt?
- Welche Prozesse sind nach der Übergabe von Marketing und Vertrieb einzuhalten?
- Wann und wie endet der Lead-Scoring-Prozess?
- Durch welche KPIs werden die Ergebnisse des Lead Scoring Prozesses gemessen?

Datengenerierung

Der erste Schritt des Lead Scorings die Generierung von historischen Daten. Hierbei werden Daten gesammelt, die sowohl alle potenziell wichtigen Leadmerkmale, also implizite und explizite Parameter, als auch die finale Disposition der Leads beinhalten (vgl. Monat 2011: 188). Diese Daten werden im Anschluss genutzt, um folgende Fragen zu beantworten (vgl. Adobe 2019: 16):

- Welche Touchpoints durchlaufen Leads, die zu SALs beziehungsweise Kunden werden?
- Welche Touchpoints durchlaufen Leads, die vom Vertrieb abgelehnt oder nicht zum MQL werden?
- Welche Eigenschaften haben MQLs?
- Welche Eigenschaften haben Leads die vom Vertrieb abgelehnt oder nie den Status MQL erreichen?
- Wie viele Touchpoints durchlaufen Leads durchschnittlich, bevor diese den Status MQL erhalten?
- Wie lange dauert es durchschnittlich, bis ein Lead zum MQL wird?

Zur Beantwortung der relevanten Fragestellungen empfiehlt es sich, folgende Daten zu generieren (vgl. Nygard und Mezei 2020: 1443):

- Lead-ID zur eindeutigen Identifikation eines Leads
- Datum der Lead-Erfassung
- Relevante explizite Lead-Daten
- Auflistung der Lead-Aktivitäten mit Zeitstempel
- Lead-Status ("SAL", "Abgelehnt", "Keine Übergabe")
- Sofern vorhanden, Datum der Leadübergabe an den Vertrieb
- Datum der letzten Lead-Aktivität

Neben den automatisch generierten Daten empfiehlt es sich, bei der Datenerhebung auch das Feedback des Vertriebsteams einzuholen. Insbesondere sollten Probleme mit den aktuell an das Team übergebenen Leads besprochen werden. So kann ermittelt werden, welche Daten aktuell häufig fehlen oder was die häufigsten Gründe für die Ablehnung von Leads sind (vgl. Adobe 2019: 16). Wie in Kapitel 3.3 beschrieben, empfiehlt es sich auch, eine Strategie zu entwickeln, um die Vollständigkeit und Qualität der Daten zu gewährleisten. Dazu können beispielsweise Datenergänzungsdienste oder das Progressive Profiling genutzt werden.

Datenaufbereitung

In mehreren der untersuchten Texte wird kritisiert, dass die Scorecard im traditionellen Lead Scoring nicht auf Daten basiert. Daher wird in diesem Schritt eine Datenanalyse durchgeführt, um ein generisches Lead-Scoring-Modell zu erstellen. Da es sich auch beim traditionellen Lead Scoring um eine Form des Data Mining handelt, werden hier die folgenden Schritte der Data Preparation beziehungsweise Datenaufbereitung aus dem CRISP-DM durchgeführt:

- Datenauswahl: Zunächst wird ausgewählt, welche Daten für die Analyse verwendet werden sollen. Diese Entscheidung basiert auf Kriterien wie der Relevanz für die Analyseziele, der Datenqualität und technischen Einschränkungen.
- Datenbereinigung: In diesem Schritt werden unvollständige Daten entweder entfernt oder durch Techniken wie Mittelwertbildung oder Medianbildung ersetzt. Außerdem werden Ausreißer entfernt und offensichtlich falsche Werte korrigiert, da diese die Qualität der Datenanalyse negativ beeinflussen können. Daten, die für die Datenanalyse nicht relevant sind, werden ebenfalls gelöscht.
- Datenkonstruktion: Nach der Datenbereinigung werden im Rahmen der Datenkonstruktion gegebenenfalls neue Datensätze oder aus bestehenden

Datensätzen abgeleitete Attribute gebildet, um die Ergebnisse der Datenanalyse zu verbessern.

- Datenintegration: Bei der Datenintegration werden Informationen aus mehreren Tabellen oder Datensätzen zusammengeführt, um eine Tabelle zu erstellen, die alle für die Datenanalyse benötigten Informationen enthält. Zur Datenintegration gehören auch Aggregationen. Aggregationen beziehen sich auf Operationen, bei denen neue Werte berechnet werden, indem Informationen aus mehreren Datensätzen und Tabellen zusammengefasst werden.
- Datenformatierung: Im Zusammenhang mit dem Data-Mining-Produkt kann es erforderlich sein, das Format oder die Struktur der Daten anzupassen. Diese Anpassungen können einfach sein, wie beispielsweise das Entfernen unzulässiger Zeichen aus Zeichenketten oder das Kürzen auf eine maximale Länge, oder komplexer, wie beispielsweise die Neuorganisation von Informationen.

Datenanalyse

Nach der Aufbereitung der Daten werden diese ausgewertet. Dazu wird zunächst eine univariate und anschließend eine bivariate Analyse durchgeführt. Bei der univariaten Analyse wird jede Variable einzeln betrachtet. Dabei können statistische Hilfsmittel wie Mittelwert, Median, Standardabweichung, Minimum, Maximum, Streuungsbreite und Histogramme verwendet werden (vgl. Cleff 2019: 26–52). Ein Beispiel aus dem Lead Scoring wäre, dass für die Anzahl der Besuche von Preisseiten pro Kontakt ein Mittelwert, eine Standardabweichung und ein Histogramm gebildet werden. Dadurch kann die Variable besser verstanden werden.

Die bivariate Analyse ist ein statistischer Ansatz, bei dem zwei unterschiedliche Variablen miteinander verglichen werden, um Beziehungen, Muster oder Korrelationen zwischen ihnen zu erkennen (vgl. Cleff 2019: 71–110). Im Lead Scoring kann beispielsweise die Korrelation zwischen der Anzahl der Besuche von Preisseiten eines Leads und dem Erreichen des MQL-Status ermittelt werden.

Entwickeln des Punktesystems

Nach der Datenanalyse kann ein datengestütztes Bepunktungsmodell erstellt werden. Die Punktevergabe erfolgt hierbei durch Experten, allerdings nutzen diese die Ergebnisse der Datenanalyse zur Validierung ihrer Einschätzungen. Damit wird dem Problem entgegengewirkt, dass traditionelle Lead-Scoring-Systeme ausschließlich auf subjektiven Experteneinschätzungen beruhen. Um das Punktemodell abzubilden, werden in einer expliziten und einer impliziten Scorecard den einzelnen Handlungen und Eigenschaften Punkte zugewiesen. Diese werden bei Erfüllung zum jeweiligen impliziten beziehungsweise expliziten Score addiert oder subtrahiert.

Wie im Kapitel 3.3 beschrieben, stellt auch der zeitliche Verlauf eine Herausforderung im Lead Scoring dar. Aktionen, die vor einem Zeitraum durchgeführt wurden, der weit über die üblichen Vertriebszyklen hinausgeht, sind keine exakten Indikatoren dafür, dass noch Kaufinteresse besteht. Eine Methode, die Anbieter von MAS und CRM-Systemen empfehlen, um ältere Aktionen mit geringerem Wert in das Lead Scoring einfließen zu lassen, sind Verfallsmodelle. Im Rahmen der Recherche wurden drei Verfallsmodelle identifiziert. Diese sind der Verfall der Punkte nach eine zuvor festgelegten Zeit, das Halbieren des Scores nach längerer Inaktivität und das Verringern des Scores nach längerer Inaktivität (vgl. ActiveCampaign o. D.; Encharge.io 2021; ConstantContact 2023):

- Verfall der Punkte nach vorgegebener Zeit: Hier wird für jede Punktevergabe festgelegt, nach welcher Zeit die vergebenen Punkte verfallen und somit vom Score des Leads abgezogen werden (vgl. ActiveCampaign o. D.).
- Halbieren der Punkte nach längerer Inaktivität: Bei dieser Verfallsmethode wird die letzte Aktivität des Kontakts geprüft. Überschreitet diese einen bestimmten Zeitraum, wird der Lead Score halbiert (vgl. ConstantContact 2023).
- Reduzierung der Punkte nach längerer Inaktivität: Auch bei dieser Verfallsmethode wird die letzte Aktivität des Kontakts überprüft. Überschreitet diese einen bestimmten Zeitraum, wird dem Score eine festgelegte Anzahl von Punkten abgezogen. Beispielsweise werden nach 30 Tagen Inaktivität 15 Punkte abgezogen, nach 60 Tagen Inaktivität 30 Punkte und nach 90 Tagen Inaktivität 50 Punkte (vgl. Encharge.io 2021).

Es ist jedoch zu beachten, dass die Verfallsmodelle zwar von verschiedenen Anbietern von MAS und CRM-Systemen empfohlen werden, jedoch noch nicht wissenschaftlich untersucht wurde, ob durch ein Verfallsmodell tatsächlich ein besseres Ergebnis erzielt wird. Die Fragen, inwiefern sich ein Verfallsmodell auf den Erfolg des Lead Scorings auswirkt, wie das richtige Verfallsmodell ausgewählt und wie es erfolgreich konfiguriert werden kann, gehen jedoch über den Rahmen dieser Forschungsarbeit hinaus.

Festlegen des Schwellenwertes

Nachdem festgelegt wurde, wie die einzelnen Parameter bewertet werden, wird der Schwellenwert festgelegt. Der Schwellenwert ist der Score, ab dem ein Lead als MQL eingestuft und an den Vertrieb übergeben wird. Der Schwellenwert wird festgelegt, indem rückwirkend ermittelt wird, welche Lead Scores vergangene SALs zum Zeitpunkt der Übergabe an den Vertrieb hatten. So kann festgelegt werden, wie hoch der Score sein muss, damit qualifizierte Leads übergeben werden. Zusätzlich werden abgelehnte und disqualifizierte Leads analysiert. Dadurch kann festgelegt werden, wie niedrig der

Schwellenwert sein darf, ohne dass schlechte oder noch nicht qualifizierte Leads weitergeleitet werden.

Anpassen des Lead-Scoring-Systems

Grundsätzlich ist zu beachten, dass mehrere Durchläufe notwendig sind, um ein robustes Lead-Scoring-Modell zu entwickeln (vgl. Hannig 2021: 250). Daher ist es sinnvoll, nach dem ersten Durchlauf den Schwellenwert, den Score und den Punkteverfall anzupassen und zu testen, ob dadurch mehr Leads korrekt vorhergesagt werden. Anschließend werden die Punktwerte und der Schwellenwert gewählt, mit denen das Modell die besten Ergebnisse erzielt.

Auswerten der Lead Scores

Bei der Auswertung der Leads werden verschiedene Pfade empfohlen:

Erfolgspfad

Auf diesem Pfad sammeln Leads Punkte, bis die zuvor festgelegten impliziten und expliziten Schwellenwerte überschritten werden. Danach werden sie zu MQLs und automatisiert an den Vertrieb übergeben. Wie in Kapitel 3.3 beschrieben, ist das Lead-Scoring-System auf eine ausreichende Menge an Verhaltensdaten angewiesen. Aus diesem Grund ist es eine Herausforderung, Leads zu identifizieren, die von Anfang an bereit für den Kauf sind. Daher werden aus der Datenanalyse Parameter ermittelt, die für ein sofortiges Kaufinteresse sprechen und dementsprechend ein beschleunigter Pfad im Lead Scoring entwickelt, auf welchem die Leads sofort an den Vertrieb weitergeleitet werden (vgl. Gooding 2022: 235).

Disqualifizierungspfad

Auf dem Disqualifizierungspfad werden Leads, die vom Vertrieb abgelehnt wurden, aus dem Lead-Scoring-System entfernt. Darüber hinaus gibt es Kontakte, bei denen von vornherein klar ist, dass es sich um Forschungsstudenten, Wettbewerber oder ähnliche Kontakte handelt. Um zu verhindern, dass solche Leads den Vertrieb erreichen, werden zunächst auf Basis der Datenanalyse Regeln definiert, um diese zu identifizieren. Anschließend werden Workflows entwickelt, um diese Leads automatisiert aus dem Lead-Scoring-System zu entfernen (vgl. Gooding 2022: 236).

Recycling-Pfad

Auf dem Recycling-Pfad werden Leads, die mit dem Vertriebsteam interagiert haben, aber noch nicht zum Kauf bereit sind, an das Marketing zurückgegeben. Auch Kontakte, die zunächst als SAL eingestuft wurden, dann aber im Vertriebsprozess verloren wurden, sollten identifiziert und an das Marketing zurückgegeben werden. Eine Möglichkeit, den Recycling-Pfad zu gestalten, besteht darin, den Vertriebsmitarbeitern die Möglichkeit zu geben, Gründe für das Nichtzustandekommen eines Geschäftsabschlusses auszuwählen. Abhängig von

diesen Gründen kann das Marketing automatisch vordefinierte Maßnahmen ergreifen, um diese Kontakte auszusortieren oder erneut durch das Marketing zu qualifizieren. So können Leads, die bereits als verloren galten, solange warmgehalten werden, bis ein konkretes Interesse vorliegt (vgl. Gooding 2022: 236).

Messen des Erfolgs

Nachdem ein Lead-Scoring-Modell in Betrieb gegangen ist, gilt es zu prüfen, wie erfolgreich dieses ist. Schließlich ist es notwendig, herauszufinden, ob das Lead Scoring tatsächlich positive Resultate bringt. Hierzu können unter anderem die in 3.3 ermittelten KPIs verwendet werden. Dazu gehören:

- Conversion Rate
- Umsatz
- Gewinn
- Kosteneinsparung
- Anzahl an Leads, die den Status MQLs erhalten
- Vertriebsproduktivität in Form des Umsatzes pro Vertriebsmitarbeiter
- Anteil der vom Vertriebsteam akzeptierten MQLs
- Erfolgsquote bei SALs
- Dauer des Vertriebszyklus
- Umsatz pro Abschluss

Aktualisieren des Lead-Scoring-Systems

Lead-Scoring-Systeme sollten mindestens quartalsweise überarbeitet werden, um sicherzustellen, dass dieses noch mit der Realität übereinstimmt. Hierzu werden folgende Aktionen durchgeführt (vgl. Kapitel 3.3):

- Prüfen der KPIs
- Analysieren der Punktestände der Leads
- Anpassen der Bepunktung durch Analyse von disqualifizierten Leads, MQLs und Ausreißern
- Einbeziehen neuer Marketingmaterialien in das Lead-Scoring-System
- Anpassen der Bepunktung und des Schwellenwertes

Im Anschluss werden die Änderungen im Lead-Scoring-Prozess dokumentiert. Um die Aktualisierung des Lead-Scoring-Systems noch effizienter zu betreiben, empfiehlt es sich außerdem einen Feedback-Mechanismus zu integrieren. Hierbei gibt das Vertriebsteam Feedback zur Qualität jedes einzelnen MQLs. In Anschluss kann dieses Feedback genutzt

werden, um das System beziehungsweise die Bepunktung zu verbessern (vgl. D'Haen und van den Poel 2013: 544–551).

6.2 Prädiktives Lead Scoring

Neben dem generischen Vorgehensmodell zum Erstellen eines traditionellen Lead-Scoring-Systems wurde auch ein Vorgehensmodell zum Erstellen eines prädiktiven Systems entwickelt. In diesem sind einige Schritte identisch, jedoch werden die Schritte "Datenanalyse", "Entwickeln des Punktesystems", "Festlegen des Schwellenwertes", "Anpassen des Lead-Scoring-Systems" und "Auswerten der Lead Scores" durch die Schritte "Modellierung", "Modell-Evaluation" und "Modell-Implementierung" ersetzt (siehe Abbildung 5). Die einzelnen Schritte des Vorgehensmodells wurden durch eine Kombination der in der Literatur identifizierten notwendigen Schritte zum Erstellen eines Lead-Scoring-Systems und der Schritte des CRISP-DM erarbeitet.

Erstellen eines Service Level Agreements

Die Erstellung eines SLA ähnelt dem Schritt "Business Understanding" des CRISP-DM (siehe Kapitel 4.1). Dabei werden die Ziele und Rahmenbedingungen des Lead Scorings zwischen Marketing und Vertrieb definiert.

Datengenerierung

Wie beim traditionellen Lead Scoring werden auch hier Daten gesammelt, die sowohl alle potenziell wichtigen impliziten und expliziten Parameter als auch Informationen zur endgültigen Disposition der Leads beinhalten.

Datenaufbereitung

Ebenso wie beim traditionellen Vorgehensmodell werden auch beim prädiktiven Vorgehensmodell die Schritte der Datenauswahl, Datenbereinigung, Datenkonstruktion, Datenintegration und Datenformatierung durchlaufen. Im Rahmen der Datenformatierung werden die Daten beim prädiktiven Scoring zusätzlich in Trainings- und Testdaten aufgeteilt und skaliert, um so ein robusteres Modell entwickeln zu können.

Modellierung

Im Modellierungsschritt werden zunächst mehrere Algorithmen ausgewählt, die im Anschluss getestet werden. Für das prädiktive Lead Scoring sind Linear Regression, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine und neuronale Netze gängige Algorithmen. Diese Algorithmen werden dann mit den zuvor aufbereiteten Trainingsdaten gespeist, um ein Modell zu entwickeln. Dabei werden unterschiedliche Hyperparameter für jeden Algorithmus getestet. Anschließend wird die Kombination aus Algorithmus und Hyperparameter-Einstellungen ausgewählt, die bei Anwendung auf die Testdaten die akkuratesten Vorhersagen liefert.

Modell-Evaluation

In der Evaluationsphase wird nach Schwachstellen im Modell und in dessen Konstruktionsprozess gesucht. Werden solche identifiziert, wird das Modell angepasst.

Modell-Implementierung

In dieser Phase wird das Modell in die MAS integriert. Außerdem werden alle Automationen erstellt, damit die vom Machine-Learning-Modell prognostizierten MQLs automatisch an das Vertriebsteam übergeben werden.

Messen des Erfolgs

Nach der Inbetriebnahme wird der Erfolg des Systems anhand von KPIs gemessen. Verschlechtern sich die KPIs, muss das Machine-Learning-Modell gegebenenfalls neu antrainiert werden.

Aktualisieren des Lead-Scoring-Systems

Ebenso wie traditionelle Lead-Scoring-Systeme müssen auch prädiktive Systeme regelmäßig aktualisiert werden. Dazu empfiehlt es sich, die Schwächen des aktuellen Modells zu identifizieren und anschließend ein neues Modell anzutrainieren. Darüber hinaus kann der Prozess der Datenaufbereitung und der Modellierung angepasst werden, um die Schwächen des aktuellen Systems zu beseitigen.

7 Anwendung des Vorgehensmodells auf die Software Mautic

In diesem Kapitel wird zunächst die MAS Mautic eingeführt. Im Anschluss wird das zuvor erstellte generische Vorgehensmodell zum Entwickeln eines traditionellen und eines prädiktiven Lead-Scoring-Systems auf Mautic angewandt. Ebenso wird untersucht, wie erweiterte Anwendungsfälle innerhalb der Software abgebildet werden können.

7.1 Einführung der Software Mautic

Bei Mautic handelt es sich um eine Open Source MAS. Der Begriff Open Source Software bezeichnet Programme, deren Quellcode öffentlich zugänglich ist und es einer globalen Entwicklergemeinschaft ermöglicht, gemeinsam an der Verbesserung des Programms zu arbeiten (vgl. Wu und Lin 2001: 33). Die Software wird von über 200.000 Unternehmen genutzt und von über 1.000 freiwilligen Helfern unterstützt (vgl. mautic.org o. D.). Wie in Kapitel 2.5 beschrieben, muss eine MAS die folgenden Aufgaben im Lead Management unterstützten und dementsprechend über dafür notwendigen Komponenten verfügen:

- Lead Generierung und Identifikation
    - Landing Pages

    - Formulare
    - Social Media Marketing
- Lead Nurturing
    - Kampagnen
    - E-Mail-Marketing
    - Social Media Marketing
    - Segmentierung und dynamische Inhalte
- Leadqualifizierung
    - Lead-Scoring-System
- Lead Routing
    - CRM-Integration
- Abschluss und darüber hinaus
    - Reports

Im Folgenden wird auf diese einzelnen Bestandteile innerhalb der Software Mautic eingegangen. Eine Gliederung der einzelnen Punkte ist dem Menü der Software (vgl. Abbildung 6) zu entnehmen.

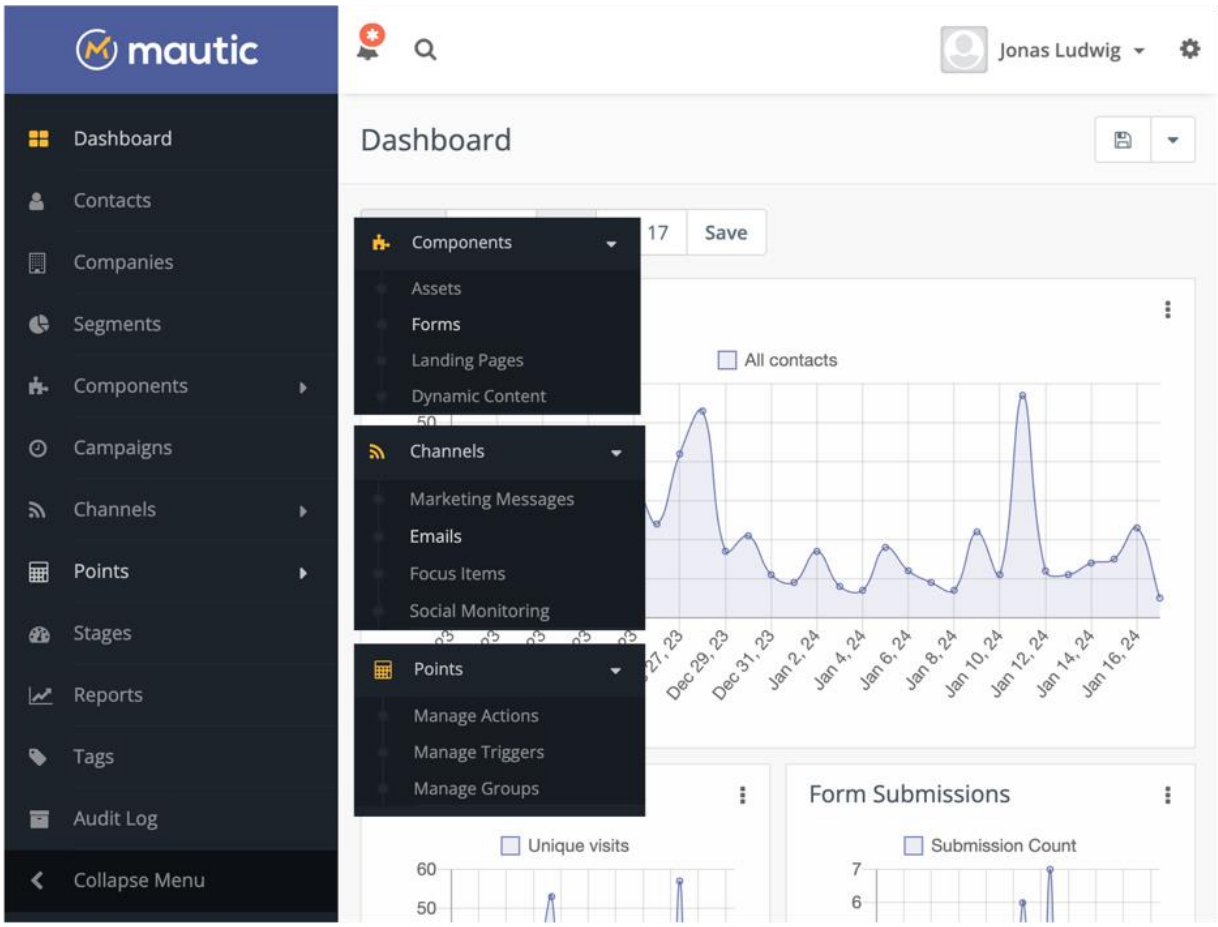

Abbildung 6: Benutzeroberfläche der Software Mautic
Quelle: Eigene Darstellung

Dashboard

Das Dashboard in Mautic ist die zentrale Anlaufstelle für Nutzer, um einen Überblick über ihre Marketingaktivitäten zu erhalten. Es zeigt wichtige Kennzahlen wie Besucherzahlen, E-Mail-Performance und andere relevante Statistiken an.

Contacts

Der Menüpunkt "Contacts" in Mautic ermöglicht die Verwaltung von Kontakten. Hier können Leads hinzugefügt, bearbeitet und nach verschiedenen Kriterien gefiltert werden.

Companies

Unternehmen können in Mautic verwendet werden, um Account Based Marketing zu betreiben. Hierzu werden Leads übergeordneten Unternehmen zugeteilt. Unter dem Punkt "Companies" können die Unternehmensinformationen verwaltet werden.

Segments

Der Punkt "Segments" ermöglicht in Mautic die gezielte Gruppierung von Kontakten aufgrund gemeinsamer Merkmale oder Verhaltensweisen. Segmente ermöglichen eine präzise Zielgruppenansprache und können für personalisierte Marketingkampagnen genutzt werden. Nutzer können Segmente nach ihren spezifischen Anforderungen erstellen und anpassen, um eine effektive Verwaltung und Analyse von Kontakten zu gewährleisten.

Components

Der Menüpunkt "Components" in Mautic fasst verschiedene Marketingelemente zusammen:

- Assets: Hier werden Dateien, wie beispielsweise PDFs abgelegt, die anschließend zum Download angeboten werden können.
- Forms: Dieser Bereich ermöglicht die Erstellung und Verwaltung von Formularen, die in Websites oder E-Mails integriert werden können, um Informationen von Kontakten zu sammeln.
- Landing Pages: Hier können über einen Drag-and-Drop-Builder Landing Pages gestaltet werden, die dazu dienen, Besucher in Leads zu konvertieren, indem sie zu einer bestimmten Aktion auffordern.
- Campaigns: Kampagnen ermöglichen das Zusammenführen einzelner Marketingaktivitäten zu umfassenden Kampagnen. Dadurch werden automatisierte Abläufe erstellt, die verschiedene Aktionen und Ereignisse kombinieren.

Channels

Channels beziehungsweise Kanäle umfassen verschiedene Kommunikationskanäle für Marketingaktivitäten:

- Marketing Messages: Marketing Messages in Mautic ermöglicht es, Inhalte zu erstellen und diese über verschiedene Kanäle wie E-Mail, SMS, Browser-Benachrichtigungen, mobile Benachrichtigungen und Tweets verfügbar zu machen. Mautic sendet die Nachrichten dann über die bevorzugten Kanäle der einzelnen Kontakte und wechselt bei Bedarf zu einem alternativen Kanal.
- Emails: Dieser Bereich ermöglicht die Erstellung, Verwaltung und Automatisierung von E-Mails über einen Drag-and-Drop-Builder.
- Focus Items: Hier werden gezielt Elemente wie Popups oder Banner erstellt, um die Aufmerksamkeit der Benutzer zu gewinnen.
- Social Monitoring: Dieser Bereich ermöglicht das Überwachen von Aktivitäten in sozialen Medien, um Einblicke in das Verhalten und die Interessen der Zielgruppe zu gewinnen.

Points

Der Menüpunkt Points ist für das Lead Scoring vorgesehen und beinhaltet drei Unterpunkte:

- Manage Groups: Unter diesem Punkt können mehrere Scores angelegt werden. Damit sind auch komplexere Anwendungsfälle wie produktbasiertes Scoring oder die Verwendung eines impliziten und eines expliziten Scores möglich.
- Manage Actions: In diesem Bereich wird das Punktesystem abgebildet, indem spezifische Aktionen ausgewählt werden und die zugehörige Punktzahl festgelegt wird. Ebenso erfolgt die Auswahl des Scores, der durch die Aktionen gesteigert oder reduziert wird.
- Manage Triggers: Hier werden die Aktionen gewählt, die ausgeführt werden, sobald ein Score einen Schwellwert überschreitet.

Stages

In Mautic ermöglichen "Stages" die Definition verschiedener Phasen in Marketingkampagnen. Dies können zum Beispiel Stufen in einem Verkaufstrichter sein, denen Leads zugeordnet werden.

Reports

In Mautic ermöglicht "Reports" die Erstellung und Anpassung von Berichten, die einen Überblick über KPIs und Marketingaktivitäten geben. Die Benutzer können Berichte nach ihren spezifischen Anforderungen konfigurieren, um die für ihre Analyse relevanten Daten und Metriken zu erhalten.

Tags

In Mautic werden "Tags" verwendet, um Kontakte einzelnen Kategorien zuzuordnen. Diese Tags ermöglichen eine effektive Segmentierung und gezielte Ansprache von Kontakten.

7.2 Traditionelles Lead Scoring in Mautic

Im Folgenden werden die Schritte des generischen Vorgehensmodells zum Erstellen eines traditionellen Lead-Scoring-Systems aus Kapitel 6.1 auf die Software Mautic angepasst.

Erstellen eines Service Level Agreement

Der erste Schritt im Vorgehen zum Entwickeln eines traditionellen Lead-Scoring-Systems in Mautic ist die Definition eines SLAs. Dieses wird gemeinsam von Vertrieb und Marketing erstellt und hält die Ziele und Rahmenbedingungen des Vorhabens fest.

Datengenerierung

Um die im Rahmen des Lead Scorings relevanten Fragen zu beantworten und ein robustes Lead-Scoring-System zu entwickeln, müssen folgende Daten generiert werden:

- Lead-ID zur eindeutigen Identifikation eines Leads
- Datum der Lead-Erfassung
- Relevante explizite Lead-Daten
- Auflistung der Lead-Aktivitäten mit Zeitstempel
- Lead-Status ("SAL", "Abgelehnt", "Keine Übergabe")
- Datum der Leadübergabe
- Datum der letzten Lead-Aktivität

All diese Daten können innerhalb von Mautic automatisiert erfasst und in der Datenbank gespeichert werden.

Datenaufbereitung

Um die Daten für die Datenanalyse vorzubereiten, werden folgende Schritte der Datenaufbereitung durchlaufen:

- Datenauswahl: Zunächst werden von den generierten Daten alle Daten ausgewählt, die analysiert werden sollen.
- Datenbereinigung: Im Rahmen der Datenbereinigung werden Datenlücken reduziert oder beseitigt und Ausreißer beziehungsweise inkorrekte Werte entfernt.
- Datenkonstruktion: Durch die Datenkonstruktion werden aus vorhandenen Datensätze neue Features gebildet, welche eine verbesserte Datenanalyse ermöglichen. Außerdem werden textbasierte Werte in numerische Werte umgewandelt, um eine effektive Analyse zu ermöglichen.
- Datenintegration: Hierbei wird aus mehreren Tabellen oder Datensätzen eine Tabelle erstellt, welche alle Daten enthält, die für das Lead Scoring relevant sind. Außerdem werden die Daten in einer sinnvollen Struktur aggregiert.

- Datenformatierung: In manchen Fällen wird zum Abschluss der Datenaufbereitung noch das Format oder die Struktur der Daten geändert.

Datenanalyse

Für die Datenanalyse können die Menüpunkte "Dashboard" und "Reports" in Mautic hilfreich sein. Es wurde jedoch festgestellt, dass die Datenanalysefunktionen nicht ausreichen, um die im Rahmen des Lead Scorings relevanten Fragestellungen zu beantworten. Zudem bietet Mautic nur wenige statistische Werkzeuge zur Auswertung der gesammelten Daten. Aus diesem Grund werden in diesem Vorgehensmodell die relevanten Daten per SQL-Befehl aus der Datenbank der Software exportiert und mit externen Werkzeugen außerhalb von Mautic ausgewertet. Im Rahmen dieser Arbeit wird dazu Python-Code verwendet, es ist aber auch möglich, externe Programme zur Datenanalyse zu verwenden. Dort werden die Daten aufbereitet und anschließend im Rahmen der univariaten und bivariaten Datenanalyse untersucht. Dabei wird insbesondere die Frage beantwortet, wie sich die Aktivitäten von akzeptierten, abgelehnten und nicht übergebenen Leads unterscheiden. Um diese Frage zu beantworten, werden für alle akzeptierten Leads die Eigenschaften und durchlaufenen Touchpoints untersucht. So kann beispielsweise festgestellt werden, wie viele Preisseiten SALs vor der Übergabe durchschnittlich besucht haben. Darüber hinaus werden Histogramme, Standardabweichungen und andere statistische Werkzeuge verwendet, um detailliertere Einblicke in den Zusammenhang zwischen dem Erreichen des SAL-Status und dem Besuch von Preisseiten und anderen Aktionen zu erhalten. Anschließend wird die gleiche Analyse für abgelehnte und nicht übergebene Leads durchgeführt, um herauszufinden, wie sich die Aktivitäten und Eigenschaften von SALs und unqualifizierten Leads unterscheiden. Weitere Fragen, die durch die Datenanalyse beantwortet werden sollen, sind folgende:

- Wie viele Touchpoints durchlaufen Leads, bevor diese zu MQLs werden?
- Wie lange dauert es durchschnittlich, bis Leads zu MQLs werden?

Entwickeln des Punktesystems

Sobald ausreichende Einblicke in die Daten gewonnen wurden, um ein realitätsnahes Punktemodell zu entwickeln, wird durch datengestützte Expertenschätzungen eine Scorecard entwickelt. Anschließend wird das Punktemodell in Mautic erstellt. Dies geschieht über den Menüpunkt "Manage Actions".

<u>Erstellen der Scores</u>

Unter dem Menüpunkt "Manage Groups" können in Mautic sogenannte Point Groups angelegt werden. Dabei handelt es sich um zusätzliche Scores, welche neben dem vordefinierten Standard-Score angelegt werden. Aus effizienzgründen wird hierbei der

Standard-Score für den expliziten Score verwendet und die Point Group "Implicit Score" für den impliziten Score angelegt (siehe Abbildung 7).

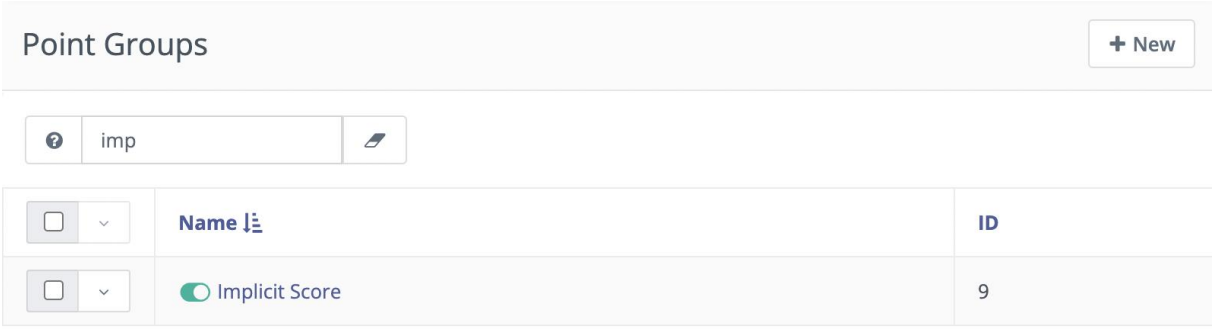

Abbildung 7: Point Groups in Mautic
Quelle: Eigene Darstellung

Festlegen der impliziten Bepunktung

Nachdem alle erforderlichen Scores erstellt wurden, können unter "Manage Actions“ Aktionen definiert werden, die den impliziten Score erhöhen oder verringern. Dazu werden die spezifischen Aktionen, die Anzahl der abzuziehenden oder hinzuzufügenden Punkte und der entsprechende Score ausgewählt. Im Beispiel in Abbildung 8 werden dem impliziten Score zehn Punkte hinzugefügt, wenn ein Kontakt das Formular "Lead Scoring Testform" ausfüllt und absendet.

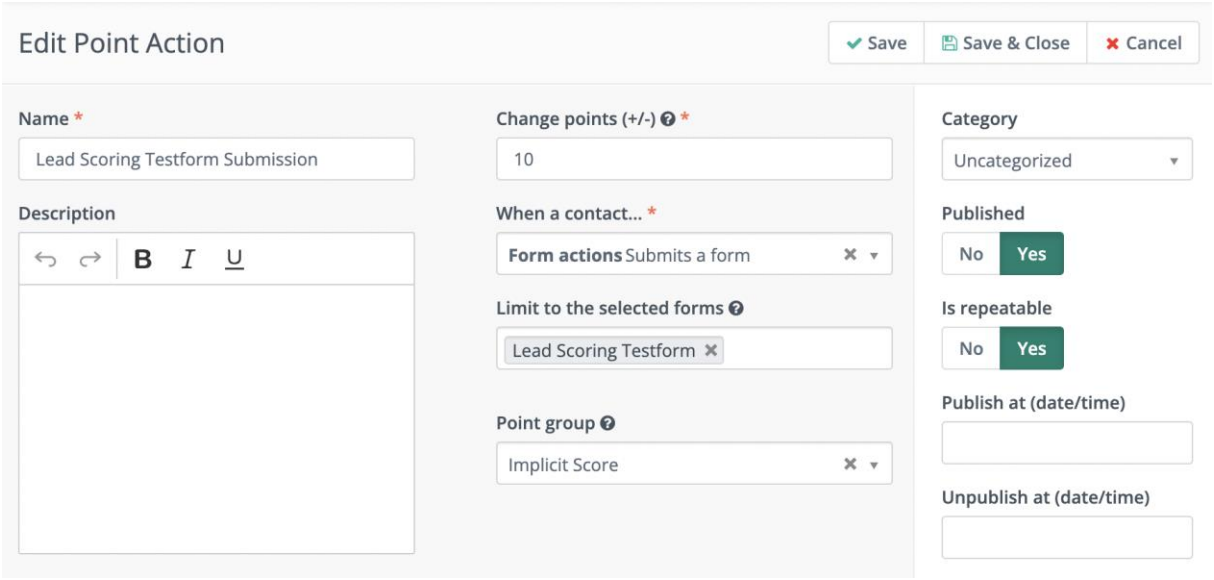

Abbildung 8: Point Actions in Mautic
Quelle: Eigene Darstellung

Unter "Manage Actions" können die folgenden Aktionen zum Erhöhen oder Verringern des impliziten Scores gewählt werden:

- Download einer Asset
- Erhalt einer E-Mail
- Öffnen einer E-Mail

- Absenden eines Formulars
- Besuch einer Landingpage oder einer URL

Folglich können nicht alle für den impliziten Score relevanten Handlungen (vgl. Kapitel 3.1) berücksichtigt werden. Beispielsweise kann die Abmeldung eines Kontaktes vom Newsletter unter "Manage Actions" nicht berücksichtigt werden. Durch den Einsatz von Kampagnen kann diese Aktion aber dennoch abgefangen werden. Dazu werden zunächst alle Kontakte, die sich abgemeldet haben, per Segmentfilter in das Segment "Do Not Contact" hinzugefügt. Anschließend wird eine Kampagne für dieses Segment gestartet. In dieser Kampagne wird jedem Kontakt im Segment mittels der Kampagnenaktion "Adjust contact points" die gewünschte Anzahl an Punkten abgezogen (siehe Abbildung 9). Außerdem wird in den Einstellungen der Kampagne festgelegt, dass die Kampagne nur einmal pro Kontakt durchlaufen werden kann, um sicherzustellen, dass nicht mehrfach Punkte für die einmalige Abmeldung aus dem Newsletter abgezogen werden.

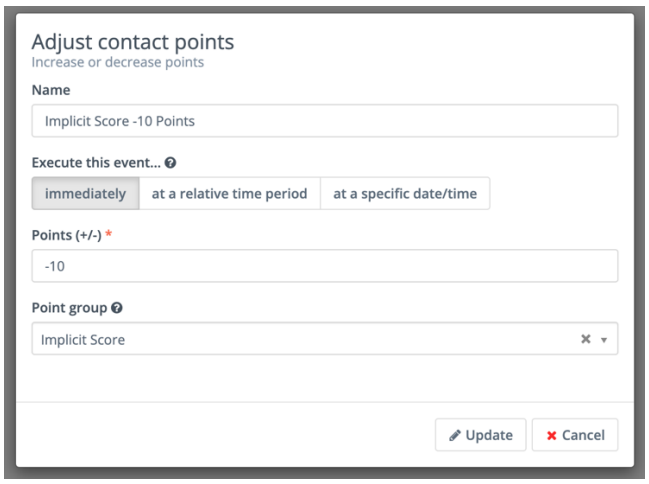

Abbildung 9: Kampagnenaktion "Adjust contact points"
Quelle: Eigene Darstellung

Festlegen der expliziten Bepunktung

Da Scores innerhalb des Menüpunkts "Manage Actions" lediglich auf Basis von Aktionen erhöht werden können, muss der explizite Score, welcher die Eigenschaften von Leads bewertet, über Kampagnen angepasst werden. Zudem muss dieser regelmäßig aktualisiert werden, um so Änderungen in den demografischen und unternehmensbezogenen Eigenschaften eines Kontaktes zu berücksichtigen.

Dazu werden zunächst alle Kontakte, die im Rahmen des Lead Scorings berücksichtigt werden, in das "Scoring Segment" aufgenommen. Auf Basis dieses Segments wird eine Kampagne zur Berechnung des expliziten Scores gestartet (vgl. Abbildung 10). Nach dem Start der Kampagne wird der explizite Score der Kontakte auf Null zurückgesetzt. Anschließend wird der neue explizite Score auf Basis der Lead-Eigenschaften berechnet.

Nachdem der explizite Lead Score auf diese Weise gebildet wurde, wird 24 Stunden gewartet, bevor dieser erneut errechnet wird.

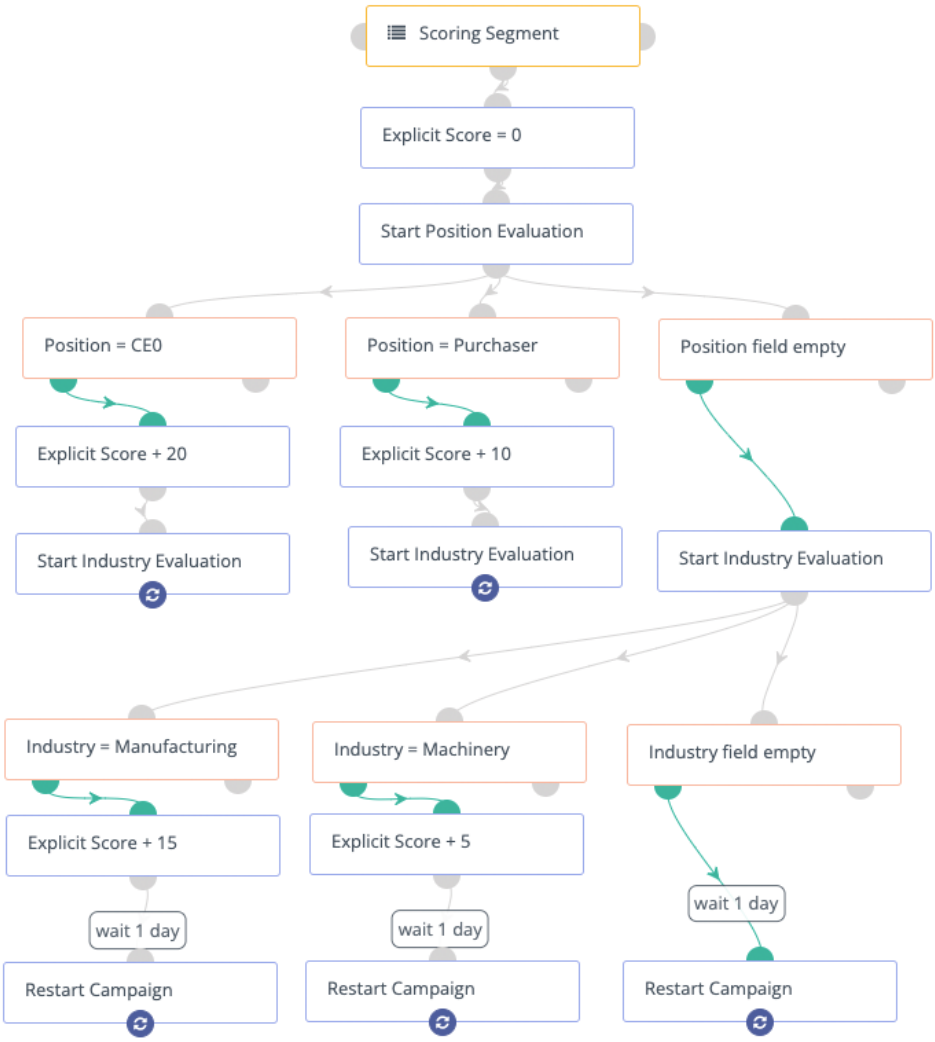

Abbildung 10: Kampagne zum Errechnen des expliziten Scores
Quelle: Eigene Darstellung

Erstellen eines Unterdrückungssegments

Um sicherzustellen, dass irrelevante Kontakte nicht an das Vertriebsteam übergeben werden, wird ein Unterdrückungssegment mit der Bezeichnung "Do Not Score" erstellt. Diesem Segment werden alle Kontakte hinzugefügt, die nicht im Lead Scoring berücksichtigt werden sollen, beispielsweise Mitarbeiter, Wettbewerber, Studenten oder Kontakte, die vom Vertrieb disqualifiziert wurden. Kontakte, die sich im Segment befinden, können später zwar noch Punkte sammeln, es werden aber keine Aktionen ausgelöst, wenn der Schwellenwert erreicht wird.

Erstellen eines Verfallsmodells

In Kapitel 6 wurden drei Verfallsmodelle ermittelt, welche sicherstellen, dass nur Leads mit aktuellem Interesse hohe implizite Socres aufweisen. Von diesen drei Modellen ist zum

derzeitigen Stand der Software Mautic nur die Reduzierung der Punkte eines Kontaktes nach längerer Inaktivität technisch möglich. Wie in Abbildung 11 dargestellt, wird dazu zunächst geprüft, wann ein Lead zuletzt aktiv war. Liegt die letzte Aktivität 30, 60 oder 90 Tage zurück, werden Punkte abgezogen. Danach wird einen Tag gewartet, damit die Kampagne nicht am selben Tag erneut gestartet wird und die Punkte für Inaktivität mehrfach abgezogen werden.

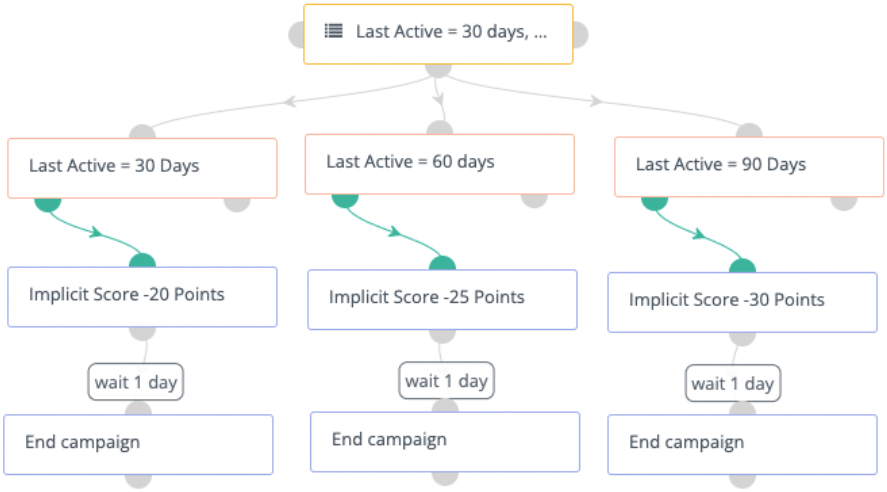

Abbildung 11: Kampagne zum Implementieren eines Verfallsmodells
Quelle: Eigene Darstellung

Erstellen einer Kampagne zum Zurücksetzen negativer Scores

Im Rahmen eines Verfallsmodells wird auch eine Kampagne zum Zurücksetzen negativer Scores entwickelt. Diese stellt sicher, dass Kontakte, denen aufgrund längerer Inaktivität Punkte abgezogen wurden, keinen negativen Score erhalten. Ansonsten besteht die Gefahr, dass ein erneut ansteigendes Interesse nicht erkannt wird, da zum Erreichen des Schwellenwertes zusätzliche Punkte gesammelt werden müssen, um den negativen Score auszugleichen.

Die Kampagne wird auf Basis eines Segments gestartet, dem Leads mit einem negativen impliziten Score hinzugefügt werden. Nach dem Start wird dem impliziten Score iterativ jeweils ein Punkt hinzugefügt und erneut geprüft, ob der Score noch negativ ist. Ist dies der Fall, wird wieder ein Punkt hinzugefügt, bis der Score den Wert Null annimmt (vgl. Abbildung 12).

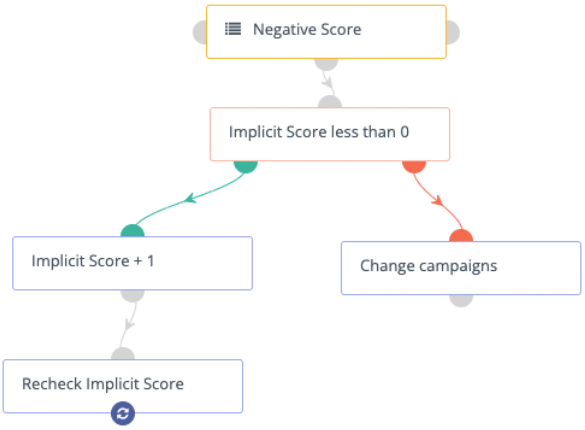

Abbildung 12: Kampagne zum Zurücksetzen negativer impliziter Scores
Quelle: Eigene Darstellung

Festlegen eines Schwellenwertes

Nachdem das Punktesystem in Mautic erstellt wurde, werden implizite und explizite Schwellenwerte ermittelt, ab welchen Leads zu MQLs werden. Die idealen Schwellwerte für den impliziten und expliziten Score werden hierbei rückwirkend mittels Python-Code ermittelt. Hierzu wird errechnet, welche Scores vergangene akzeptierte und abgelehnte Leads zum Zeitpunkt der Übergabe hatten. Zudem wird festgestellt, welchen Score Leads, die nicht übergeben wurden, zum Zeitpunkt ihrer letzten Aktivität hatten. Nach einer Analyse werden die impliziten und expliziten Schwellenwerte gewählt, für welche die höchste Anzahl an Leads korrekt vorhergesagt wird.

Anpassen des Lead-Scoring-Systems

Nach dem Erstellen des Modells wird dieses iterativ angepasst. Hierzu werden die Schwellenwerte, die Bepunktung und gegebenenfalls das Verfallsmodell angepasst und rückwirkend ermittelt, ob frühere MQLs präziser vorhergesagt werden. Nach mehreren Durchläufen werden die Parameter gewählt, mit welchen das Modell die besten Ergebnisse erzielt.

Auswertung der Lead Scores

Nachdem das Lead-Scoring-System fertiggestellt ist, werden Leads, welche den Schwellenwert überschreiten, ausgewertet. Hierzu werden die folgenden Schritte durchlaufen.

<u>Leadübergabe</u>

Sobald ein Lead die impliziten und expliziten Schwellenwerte überschreitet, wird dieser an das Vertriebsteam übergeben. Dazu wird zunächst ein Startsegment angelegt, in das Leads gelangen, die beide Schwellenwerte überschritten haben. Darauf basierend wird eine Kampagne zur Leadübergabe ausgelöst. Dabei wird das zuständige Vertriebsteam per E-Mail darüber informiert, dass ein neuer MQL vorliegt. Das für den Kontakt zuständige

Vertriebsteam wird über eine Synchronisation mit der CRM-Software ermittelt. Dabei wird in der CRM-Software anhand von Merkmalen der einzelnen Leads wie beispielsweise Region und Branche automatisiert ermittelt, welches Vertriebsteam zuständig ist. Durch die Integration mit der CRM-Software können zudem die während des Lead Scorings gesammelten Daten in die CRM-Software übertragen werden, sodass diese nach der Übergabe dem Vertriebsteam zur Verfügung stehen. Zudem werden die Kontakte mit dem Tag "Routed" versehen und automatisch der Feedback Loop Kampagne hinzugefügt (vgl. Abbildung 13).

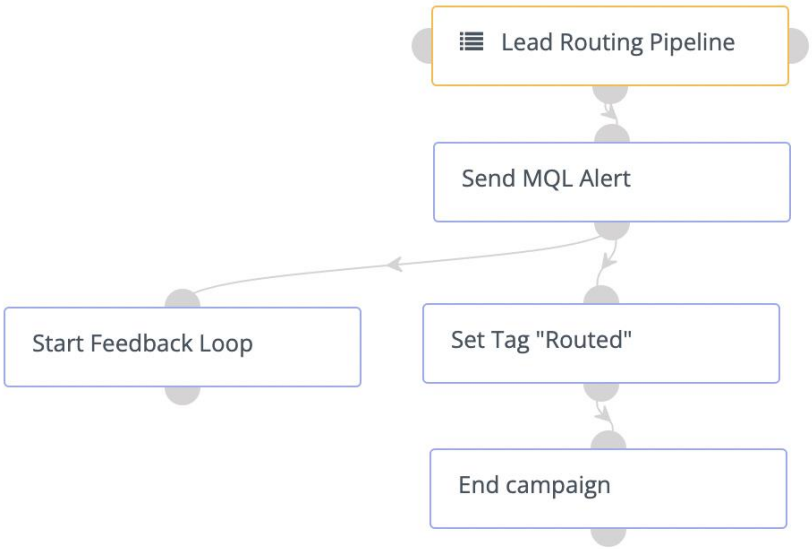

Abbildung 13: Kampagne zur Übergabe von Leads
Quelle: Eigene Darstellung

Es gilt zu beachten, dass neben der Übergabe an das Vertriebsteam in Mautic alternative Aktionen beim Erreichen des Schwellenwertes ausgeführt werden könne. Einige der in Mautic verfügbaren Aktionen sind folgende:

- Senden einer E-Mail an den Kontakt
- Entfernen oder Hinzufügen des Kontakts aus Kampagnen
- Entfernen oder Hinzufügen des Kontakts aus Segmenten
- Entfernen oder Hinzufügen von Tags
- Übergabe des Kontaktes an eine Integration

Einholen von Vertriebsfeedback

Nachdem die Leads an den Vertrieb übergeben wurden, wird das Feedback der Vertriebsmitarbeiter zur Qualität der einzelnen Leads eingeholt. Dieses Feedback unterstützt die Bildung von KPIs und ermöglicht gleichzeitig die Optimierung des Lead-Scoring-Systems. Vertriebsmitarbeiter können angeben, ob ein Lead akzeptiert wird und den Status SAL erhält, ob der Lead noch nicht bereit für den Vertriebsprozess ist und daher recycelt werden soll, oder ob der Lead abgelehnt wird. Es empfiehlt sich, einen Feedback-Prozess festzulegen und in das CRM-System zu integrieren, das mit der Marketing-Automatisierungssoftware

synchronisiert ist. Alternativ kann das Feedback auch durch eine Feedback-Loop-Kampagne in Mautic generiert werden, auf die in Anhang 1 näher eingegangen wird.

Auswertung des Vertriebsfeedbacks

Sobald das Vertriebsfeedback verfügbar ist, wird der Kontakt einer Kampagne zur Auswertung des Feedbacks hinzugefügt. Hierbei gibt es drei Pfade (siehe Abbildung 14):

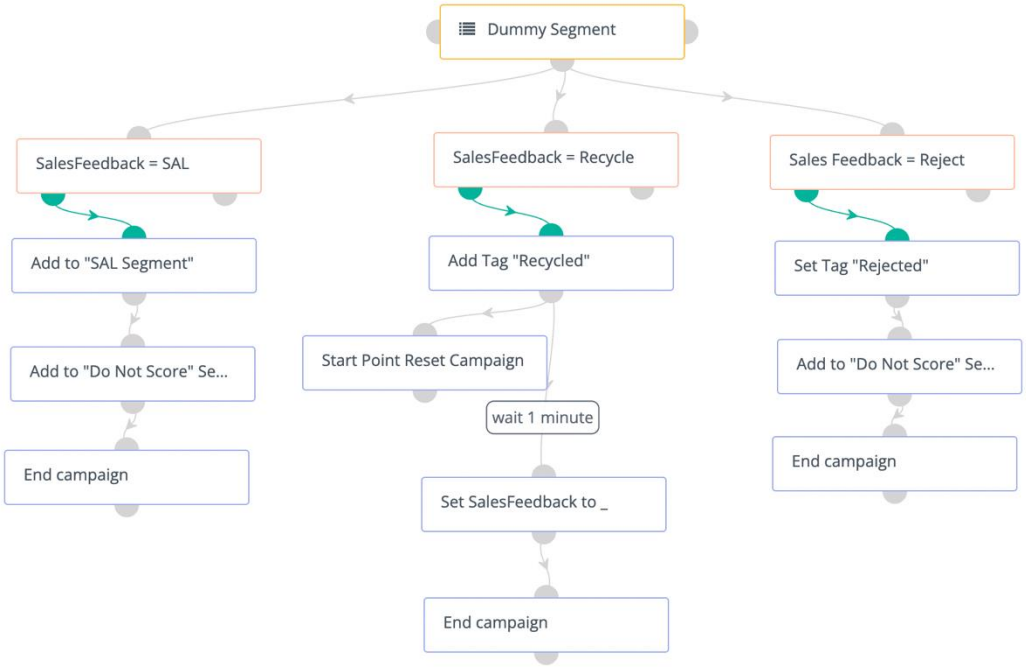

Abbildung 14: Kampagne zur Feedbackauswertung
Quelle: Eigene Darstellung

- Erfolgspfad: Der Erfolgspfad wird beschritten, wenn das Feedback "SAL" gegeben wird. In diesem Fall wird der Lead zunächst dem "Do Not Score" Segment hinzugefügt, um zu vermeiden, dass er erneut übergeben wird. Außerdem wird der Lead einem Segment für SALs hinzugefügt. Da nicht jeder SAL zum Kunden konvertiert, wird eine weitere Kampagne entwickelt, welche SALs, die im Vertriebsprozess nicht konvertiert werden können, wieder in den Lead-Scoring-Prozess integriert. Hierzu wird in Mautic ein Feld namens "Reactivate_SAL" angelegt, das mit der CRM-Software synchronisiert wird. Dieses Feld kann die Werte "Ja" oder "Nein" annehmen. Wenn es nun von einem Vertriebsmitarbeiter auf "Ja" gesetzt wird, wird eine Kampagne gestartet, um die SALs wieder in den Lead Scoring Prozess zu integrieren. Dem Kontakt wird dabei zunächst der Tag "Reactivated_SAL" hinzugefügt. Anschließend wird der implizite Score auf einen zuvor festgelegten Wert zurückgesetzt. Auch das Feedback vom Vertrieb wird zurückgesetzt. Zusätzlich wird der Lead aus dem "SAL-Segment" und dem "Do Not Score-Segment" entfernt (siehe Abbildung 15).

- Recycling-Pfad: Wird als Feedback "Recycle" abgegeben, ist der Lead momentan nicht bereit für die Bearbeitung durch das Vertriebsteam. In diesem Fall wird der Lead mit dem Tag "Recycled" markiert und an das Marketing zurückgegeben. Zudem werden die Punkte des Leads auf einen zuvor definierten Wert zurückgesetzt und der Inhalt des Feldes "SalesFeedback" wird geleert.
- Disqualifizierungspfad: Wird das Feedback "Reject" abgegeben, so wird der Lead vom Vertrieb vollständig abgelehnt. Abgelehnte Leads werden dem "Do Not Score Segment" hinzugefügt und erhalten den Tag "Rejected".

Abbildung 15: Kampagne zum Reaktivieren von Leads
Quelle: Eigene Darstellung

Messung des Erfolgs

Nachdem das System in Betrieb genommen wurde, wird mittels KPIs der Erfolg gemessen. Verschlechtern sich die Kennzahlen, so kann dies ein Zeichen dafür sein, dass das Lead-Scoring-System aktualisiert oder verbessert werden muss.

Aktualisierung des Lead-Scoring-Modells

Mindestens alle drei Monate wird das Lead-Scoring-Modell aktualisiert. Hierfür werden die Schritte zur Datenaufbereitung und Datenanalyse erneut durchgeführt und das System entsprechend aktualisiert.

7.3 Prädiktives Lead Scoring in Mautic

In diesem Kapitel werden die Schritte des generischen Vorgehensmodells zum Erstellen eines prädiktiven Lead-Scoring-Systems aus Kapitel 6.2 auf die Software Mautic angepasst.

Erstellen eines Service Level Agreements

Der erste Schritt im Vorgehen zum Entwickeln eines prädiktiven Lead-Scoring-Systems in Mautic ist die Definition eines SLAs. Hierbei werden die Ziele und Rahmenbedingungen des Lead-Scoring-Systems zwischen Marketing und Vertrieb definiert.

Datengenerierung

In diesem Schritt wird sichergestellt, dass in der Mautic Datenbank alle benötigten Daten zur Verfügung stehen. Dabei handelt es sich um folgende Daten:

- Lead-ID zur eindeutigen Identifikation eines Leads
- Datum der Lead-Erfassung
- Relevante explizite Lead-Daten
- Auflistung der Lead-Aktivitäten mit Zeitstempel
- Lead-Status ("SAL", "Abgelehnt", "Keine Übergabe")
- Datum der Leadübergabe
- Datum der letzten Lead-Aktivität

Datenaufbereitung

Um die Daten für die Entwicklung eines Modells des maschinellen Lernens aufzubereiten, werden die Daten per CSV-Export aus Mautic exportiert und dort mittels Python-Code aufbereitet. Dabei werden die Schritte Datenauswahl, Datenbereinigung, Datenkonstruktion, Datenintegration und Datenformatierung durchlaufen.

Modellierung

Nach der Datenaufbereitung wird ein robustes Machine-Learning-Modell entwickelt. Vor der eigentlichen Modellentwicklung werden zunächst verschiedene Modelle beziehungsweise Machine-Learning-Algorithmen ausgewählt, die im Rahmen des Lead Scorings getestet werden. Im Kontext des prädiktiven Lead Scorings werden hierbei vier Algorithmen verwendet, die in der untersuchten Literatur zum Thema prädiktives Lead Scoring vermehrt Anwendung finden. Es handelt sich dabei um die Algorithmen Logistic Regression, Decision Tree, Random Forest und Support Vector Machine. Zur Modellentwicklung wird mit den

Algorithmen eine Grid-Search durchgeführt. Das bedeutet, dass das Modell mit verschiedenen Kombinationen der definierten Hyperparameter trainiert und getestet wird, um die besten Einstellungen zu finden. Nachdem die Grid-Search für jedes Modell abgeschlossen ist, wird das beste Modell mit den besten Hyperparametern für die spätere Verwendung gespeichert, sodass es zukünftig zur Vorhersage neuer Daten verwendet werden kann, ohne erneut ein Modell antrainieren zu müssen. Das beste Modell wird hier durch die Genauigkeit der Modelle berechnet, die angibt, wie viel Prozent der Vorhersagen eines Modells korrekt sind.

Modell-Evaluation

In dieser Evaluationsphase wird nach Schwachstellen im Modell und in seinem Entwicklungsprozess gesucht. Außerdem wird bewertet, ob das Modell die im SLA festgelegten Ziele erfüllt. Der Prozess der Bereitstellung des Modells wird ebenfalls geplant.

Modell-Implementierung

Im Folgenden werden die einzelnen Schritte zur Anwendung des Modells auf Mautic beschrieben:

Export der Lead-Daten aus Mautic

Im ersten Schritt der Modellanwendung werden alle Leads, die nicht konvertiert sind und sich nicht im Unterdrückungssegment befinden, als CSV aus Mautic exportiert.

Datenaufbereitung

Im Anschluss werden die Lead-Daten wie bereits im Schritt der Datenaufbereitung durch den dabei erstellten Python-Code vorbereitet.

Erstellen der Vorhersagen

Anschließend wird das zuvor gespeicherte maschinelle Lernmodell geladen, um Vorhersagen für jeden Lead zu treffen. Die Vorhersagen werden zusammen mit der entsprechenden Lead-ID als CSV-Datei gespeichert.

Übertragen der Modellvorhersagen in Mautic

Nachdem die Vorhersagen getroffen wurden, wird die CSV-Datei mit den Modellvorhersagen in Mautic importiert. Hierzu wird beim Import die Spalte lead_id der Contact ID in Mautic zugeordnet und die Spalte mit den Vorhersagen einem zuvor erstellten Datenbankfeld mit dem Namen "Scoring Predictions" (siehe Abbildung 16).

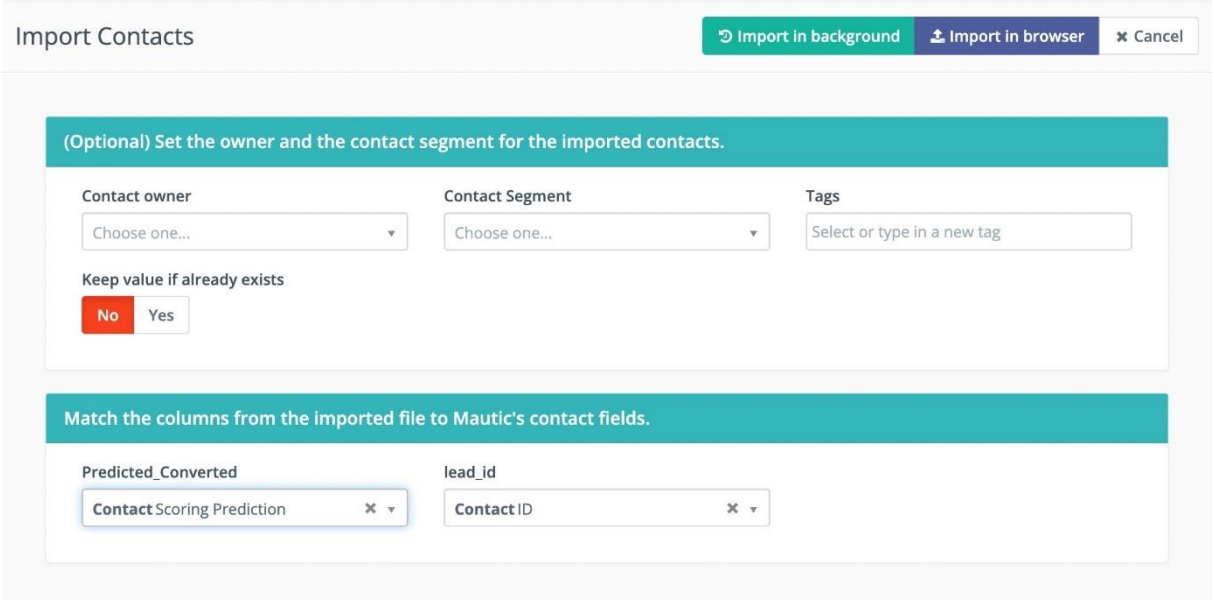

Abbildung 16: Import der Vorhersagen eines prädiktiven Lead-Scoring-Modells in Mautic
Quelle: Eigene Darstellung

Leadübergabe

Kontakte, die durch das maschinelle Lernmodell als MQLs klassifiziert wurden, werden anschließend automatisch per Kampagne an das Vertriebsteam weitergeleitet.

Auswerten der Leads

Sobald ein Lead übergeben wurde, startet der Mechanismus zur Einholung von Feedback zur Qualität des übergebenen Leads seitens des Vertriebsteams. Dieser Mechanismus kann entweder im CRM-System oder in der MAS (siehe Anhang 1) implementiert werden.

Anschließend wird eine Kampagne zur Auswertung des Feedbacks gestartet. Hier wird je nach Feedback des Vertriebs der SAL-Pfad, der Recycling-Pfad oder der Disqualifizierungspfad gewählt. Für den SAL-Pfad wird zusätzlich eine Kampagne angelegt, über die SALs, die nicht als Kunden gewonnen werden konnten, wieder in den Lead-Scoring-Prozess für die entsprechende Produktgruppe integriert werden können.

Um sicherzustellen, dass recycelte und reaktivierte Leads nicht sofort wieder als MQLs prognostiziert werden, empfiehlt es sich, vor dem Recycling oder der Reaktivierung einen zuvor festgelegten Zeitraum abzuwarten. Dadurch ändert sich deren Aktivitätsprofil, wodurch das Machine-Learning-Modell aktualisierte Daten für die Lead-Bewertung erhält.

Messen des Erfolgs

Nach der Inbetriebnahme wird der Erfolg des Systems anhand von KPIs gemessen. Verschlechtern sich die KPIs, so muss das Machine-Learning-Modell möglicherweise neu antrainiert werden.

Aktualisieren des Lead-Scoring-Systems

Genauso wie traditionelle Lead-Scoring-Systeme müssen auch prädiktive Lead-Scoring-Modelle regelmäßig aktualisiert werden. Hierzu empfiehlt es sich, Schwächen des aktuellen Modells zu ermitteln und im Anschluss ein neues Modell anzutrainieren. Hierbei kann gegebenenfalls ein abgeänderter Prozess der Datenaufbereitung- und Modellierung verwendet werden, um die Schwächen des alten Modells zu beseitigen.

7.4 Produktbasiertes Lead Scoring in Mautic

Im Zuge dieser Forschungsarbeit wurde ebenfalls der Anwendungsfall des produktbasierten Lead Scorings in Mautic implementiert. Hierbei ist zu beachten, dass für jede Produktgruppe ein eigenständiges Lead-Scoring-System entwickelt werden muss. Die Erstellung dieser individuellen Systeme kann je nach präferiertem Ansatz entweder den Schritten zur Gestaltung eines traditionellen Lead-Scoring-Systems (sehe Kapitel 7.2) oder den Schritten zur Entwicklung eines prädiktiven Lead-Scoring-Systems (siehe Kapitel 7.3) folgen. Da mehrere Produktbereiche in das Lead Scoring integriert werden, müssen hier möglicherweise mehrere SLAs mit unterschiedlichen Vertriebsteams erstellt werden.

7.5 Account Based Scoring im Mautic

Neben dem traditionellen Lead Scoring und dem produktbasierten Scoring wurde auch das ABS in Mautic untersucht. Das traditionelle Punktesystem kann hier mithilfe von Kampagnen abgebildet werden (siehe Abbildung 17). Beim Testen der ABS-Funktionalitäten der Software wurde jedoch festgestellt, dass es innerhalb von Mautic keine Möglichkeit gibt, Aktionen auf Basis des Scores eines Accounts auszuführen. Auch beim prädiktiven ABS ist es zwar möglich, alle einem Unternehmen zugeordneten Eigenschaften und Aktionen zu exportieren, auszuwerten und die Modellvorhersagen zu importieren, allerdings liegt auch hier das Problem darin, dass es nicht möglich ist, basierend auf den Vorhersagen des Modells Aktionen auf Unternehmensebene zu starten. Daher reichen die ABS-Funktionalitäten in Mautic nicht aus, um einen automatisierten Scoring-Prozess zu ermöglichen.

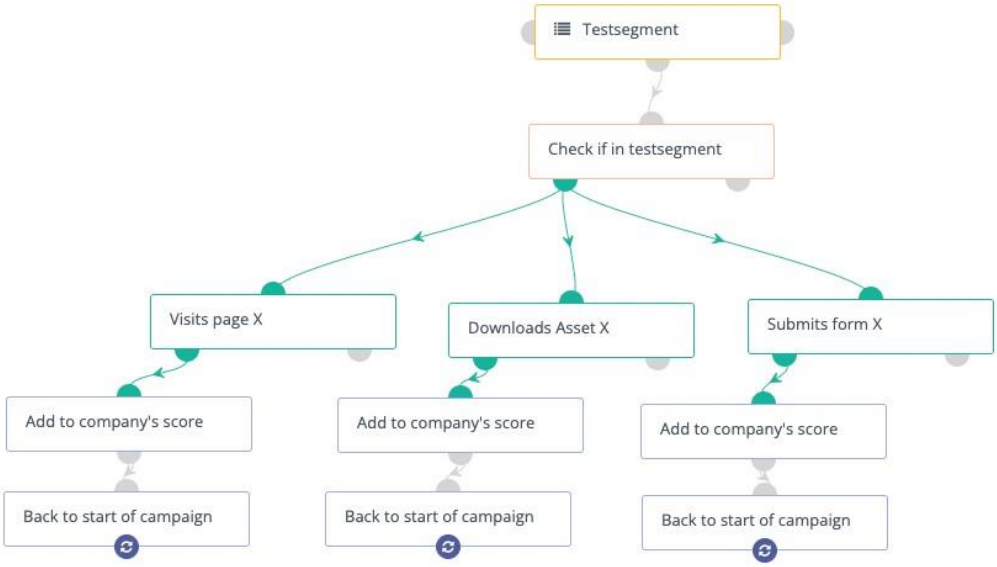

Abbildung 17: Kampagne zum Errechnen des impliziten Scores im Account Based Scoring
Quelle: Eigene Darstellung

8 Praktische Umsetzung und Ergebnisse des Lead Scorings in Mautic

In diesem Teil der Arbeit wird die Anwendung des zuvor ermittelten Vorgehensmodells zur Entwicklung eines Lead-Scoring-Systems auf einen realen Onlineshop angewandt, der die MAS Mautic einsetzt. Dabei werden sowohl ein traditionelles als auch ein prädiktives Lead-Scoring-System entwickelt und deren Resultate am Beispiel des Online-Shops verglichen.

8.1 Traditionelles Lead Scoring

Erstellen eines Service Level Agreement

Da kein Kontakt zum Marketing- und Vertriebsteam des Onlineshops besteht, wurde die Erstellung eines SLA im Rahmen des Praxisbeispiels nicht durchgeführt.

Datengenerierung

Als Datenbasis für das Lead Scoring werden die Daten verwendet, die sich bereits innerhalb der Mautic Datenbank des Onlineshops befinden. Hierbei stehen alle Websitebesuche inklusive Zeitstempel und, falls sich ein Kontakt vom Newsletter abgemeldet hat, das Datum der Newsletterabmeldung zur Verfügung. Außerdem kann aus den Segmentzugehörigkeiten der Kontakte abgeleitet werden, ob es sich um konvertierte oder nicht-konvertierte Leads handelt.

Datenaufbereitung

Nachdem die Daten generiert wurden, werden diese innerhalb von Python aufbereitet. Hierzu wird der Python-Code aus Anhang 2 verwendet. Auf die einzelnen Schritte der Datenaufbereitung wird im Folgenden eingegangen.

Datenauswahl

In diesem Schritt werden zunächst mithilfe von zwei SQL-Statements (siehe Anhang 3 und 4) die Daten konvertierter und nicht-konvertierter Leads als CSV-Dateien exportiert. Eine anonymisierte Version der Tabelle für konvertierte Leads ist in Tabelle 11 zu sehen und enthält folgende Spalten:

- lead_id: Die lead_id für dient zur eindeutigen Identifizierung eines Kontakts.
- end_date: Das end_date steht für den Zeitpunkt, zu welchem ein Kontakt zum Kunden konvertiert ist oder für den Zeitpunkt der letzten Aktivität bei nicht-konvertierten Leads.
- url: In der URL-Spalte wird für jeden Aktivitätseintrag die besuchte Webseite angegeben.
- date_hit: Das date_hit ist der Zeitstempel zum Besuch einer Website.
- start_date: Das start_date beschreibt den Zeitpunkt, zu welchem ein Lead in Mautic erfasst wurde.
- confirmed_optout_date: Hier wird das Datum der Abmeldung vom Newsletter gespeichert, sofern eine Abmeldung erfolgte.

lead_id	end_date	url	date_hit	start_date	confirmed_optout_date
84654	2021-09-08 06:10:04	https://www.mywebsite.com/.....	2022-01-01 02:40:14	2021-03-01 22:22:37	
84654	2021-09-08 06:10:05	https://www.mywebsite.com/.....	2022-01-01 02:40:15	2021-03-01 22:22:38	
84654	2021-09-08 06:10:06	https://www.mywebsite.com/.....	2022-01-01 02:40:16	2021-03-01 22:22:39	
84654	2021-09-08 06:10:07	https://www.mywebsite.com/.....	2022-01-01 02:40:17	2021-03-01 22:22:40	
84654	2021-09-08 06:10:08	https://www.mywebsite.com/.....	2022-01-01 02:40:18	2021-03-01 22:22:41	
84654	2021-09-08 06:10:09	https://www.mywebsite.com/.....	2022-01-01 02:40:19	2021-03-01 22:22:42	
21544	2021-09-08 06:10:10	https://www.mywebsite.com/.....	2022-01-01 02:40:20	2021-03-01 22:22:43	
21544	2021-09-08 06:10:11	https://www.mywebsite.com/.....	2022-01-01 02:40:21	2021-03-01 22:22:44	
21544	2021-09-08 06:10:12	https://www.mywebsite.com/.....	2022-01-01 02:40:22	2021-03-01 22:22:45	
21544	2021-09-08 06:10:13	https://www.mywebsite.com/.....	2022-01-01 02:40:23	2021-03-01 22:22:46	
21544	2021-09-08 06:10:14	https://www.mywebsite.com/.....	2022-01-01 02:40:24	2021-03-01 22:22:47	
84587	2021-09-08 06:10:15	https://www.mywebsite.com/.....	2022-01-01 02:40:25	2021-03-01 22:22:48	
84587	2021-09-08 06:10:16	https://www.mywebsite.com/.....	2022-01-01 02:40:26	2021-03-01 22:22:49	
84587	2021-09-08 06:10:17	https://www.mywebsite.com/.....	2022-01-01 02:40:27	2021-03-01 22:22:50	
84587	2021-09-08 06:10:18	https://www.mywebsite.com/.....	2022-01-01 02:40:28	2021-03-01 22:22:51	
84587	2021-09-08 06:10:19	https://www.mywebsite.com/.....	2022-01-01 02:40:29	2021-03-01 22:22:52	
84587	2021-09-08 06:10:20	https://www.mywebsite.com/.....	2022-01-01 02:40:30	2021-03-01 22:22:53	
84587	2021-09-08 06:10:21	https://www.mywebsite.com/.....	2022-01-01 02:40:31	2021-03-01 22:22:54	
84587	2021-09-08 06:10:22	https://www.mywebsite.com/.....	2022-01-01 02:40:32	2021-03-01 22:22:55	
84588	2021-09-08 06:10:23	https://www.mywebsite.com/.....	2022-01-01 02:40:33	2021-03-01 22:22:56	2022-01-01 02:40:35

Tabelle 11: DataFrame konvertierter Leads
Quelle: Eigene Darstellung

Nach dem Einlesen der Daten in Python wird zunächst in beiden Tabellen beziehungsweise DataFrames (DFs) eine Spalte mit dem Namen "Coverted" hinzugefügt. Für die Einträge aus

der Tabelle der konvertierten Leads wird hier eine "1" ergänzt und für nicht-konvertierte Leads eine "0". Anschließend werden die beiden Tabellen zusammengeführt.

Datenbereinigung

Da im Anwendungsbeispiel der Fokus auf der Vorhersage der Konversion zu Neukunden liegt, sind Aktionen, die nach dem ersten Kauf stattgefunden haben, nicht relevant. Daher werden bei Kunden Webseitenbesuche und Abmeldungen vom Newsletter, die nach dem Kaufdatum liegen, entfernt.

Datenkonstruktion

Um eine aussagekräftige Datenanalyse zu ermöglichen, werden in diesem Schritt neue Daten konstruiert. Die Spalte confirmed_optout_date enthält entweder ein Datum, falls sich ein Lead vom Newsletter abgemeldet hat, oder ist leer, falls sich ein Lead nicht abgemeldet hat. Im Rahmen der Kodierung werden hierbei alle Daten in eine "1" umgewandelt und alle leeren Werte in eine "0". Außerdem wird die Spalte "salescycle_duration" erstellt, in welcher die Dauer für jeden Lead ermittelt wird, bis dieser den Status MQL erhielt. Hierfür wird die Differenz zwischen dem start_date und dem end_date berechnet. Für nicht-konvertierte Leads wird folglich der Zeitraum zwischen der Erfassung und der letzten Aktivität berechnet.

Datenintegration

Im Rahmen der Analyse der URL-Spalte wurde festgestellt, dass von den Kontakten insgesamt 2327 unterschiedliche Webseiten besucht wurden. Um die Komplexität im Lead Scoring zu reduzieren und eine akkurate Auswertung zu garantieren, werden die besuchten URLs verschiedenen URL-Kategorien zugeordnet. Hierzu wurden nach einer Analyse der Werte in der URL-Spalte die Kategorien "Basket", "Browsing 1", "Browsing 2", "Browsing 3", "Checkout", "Checkout-Registration", "Contact", "Faq", "Account", "Registration", "Newsletter Registration", "Price", "Product" und "Other Page" erstellt. Die Einteilung der URL-Spalte in URL-Kategorien erfolgt über Schlüsselwörter. Sobald jede URL durch eine Kategorie ersetzt wurde, wird in einem neuen DF für jede URL-Kategorie eine Spalte erstellt, in der festgehalten wird, wie oft ein Lead die URL-Kategorie besucht hat. Außerdem wird eine weitere Spalte erstellt, welche die gesamten Websitebesuche pro Lead aggregiert.

Im Anschluss werden die gesamten Websitebesuche, die Websitebesuche pro URL-Kategorie, die "salescycle_duration", die Spalte "Converted" und der Status zur Newsletterabmeldung gemeinsam mit der Lead-ID im DF "Leadübersicht" zusammengeführt, dessen erste Zeilen Tabelle 12 zu entnehmen sind.

lead_id	total_pagehits	Basket	Browsing 1	Browsing 2	Browsing 3	Checkout	Checkout-Registration	Contact	Faq	Account	Newsletter Registration	Other Page	Price	Product	Registration	salescycle_duration	Converted	opted_out
1	6	0	0	0	0	0	0	0	0	0	0	0	0	6	0	1018,62	0	0
2	20	1	0	0	0	2	0	0	1	0	0	2	0	14	0	811,45	1	0
3	4	0	0	0	0	0	0	0	0	0	0	0	0	4	0	796,98	0	0
4	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	796,96	0	0
5	12	2	0	0	0	4	0	0	0	0	0	0	0	6	0	796,95	0	1
6	4	0	0	0	0	0	0	0	0	0	0	0	0	4	0	796,94	0	0
7	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	796,94	0	0
8	4	0	0	0	0	0	0	0	0	0	0	0	0	4	0	796,89	0	0
9	765	44	6	37	10	77	13	6	4	61	0	84	13	406	4	796,88	0	0
10	14	0	0	1	0	0	0	0	0	2	0	6	0	5	0	796,88	0	0
11	3	0	0	0	0	0	0	0	0	0	0	0	0	2	1	796,87	0	0
12	32	0	0	0	0	0	0	0	0	0	0	0	0	32	0	796,87	0	0
13	8	0	0	0	0	0	0	0	0	0	0	0	0	8	0	796,85	0	0
14	117	7	0	0	7	22	6	0	0	11	0	21	0	43	0	796,84	0	0
15	25	2	0	0	0	0	0	0	0	0	0	0	0	23	0	796,84	0	0
16	4	0	0	0	0	0	0	0	0	0	0	0	0	4	0	796,84	0	0
17	4	0	0	0	0	0	0	0	0	0	0	0	0	4	0	796,83	0	0
18	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	796,83	0	0
19	52	7	0	6	0	9	1	0	0	4	1	4	1	19	0	796,82	0	0
20	80	9	0	1	0	25	0	0	0	3	0	4	0	38	0	796,81	0	0
21	15	0	0	0	0	0	0	0	0	0	0	0	0	15	0	796,8	0	0
22	10	0	0	0	6	0	0	0	0	0	0	1	1	2	0	796,8	0	0
23	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	796,79	0	0
24	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	796,79	0	0
25	4	0	0	0	0	0	0	0	0	0	4	0	0	0	0	796,78	0	0
26	41	2	2	2	0	2	0	0	0	1	0	0	1	31	0	796,77	0	0

Tabelle 12: Leadübersicht DataFrame
Quelle: Eigene Darstellung

Im Rahmen der Datenbereinigung werden aus dem Leadübersicht DF im Anschluss Ausreißer entfernt, um zu vermeiden, dass sie die Analyse beeinflussen. Hierbei handelt es sich um Leads, welche eine übermäßig hohe Aktivität aufweisen und mehr als 300 Websitebesuche durchgeführt haben. Zusätzlich werden Leads mit weniger als drei Websitebesuchen gelöscht, da deren Wert im Rahmen der Datenanalyse als gering eingeschätzt wird.

Datenformatierung

Eine Formatierung der Daten ist für diesen Anwendungsfall nicht erforderlich.

Datenanalyse

Zur Analyse der Daten wird der Leadübersicht DF in zwei DFs für konvertierte und nicht-konvertierte Leads aufgeteilt, um diese im Rahmen der Analyse separat zu untersuchen.

Für jede Spalte, abgesehen von der Spalte "opted_out" in den zwei DFs werden der Durchschnitt, die Standardabweichung, der Median und das Maximum berechnet. Die daraus resultierenden Werte werden in einer Statistiktabelle zusammengefasst. Außerdem wird über die Spalte "opted_out" der Prozentsatz an Kontakten ermittelt, die sich von Newslettern abgemeldet haben, sowohl für konvertierte als auch für nicht-konvertierte Leads. Dieser wird zur Statistiktabelle hinzugefügt (siehe Tabelle 13).

		Converted	Unconverted
Basket	**Average**	1,67	0,78
	Stdev	3,77	2,81
	Median	0	0
	Max	42	42
Browsing 1	**Average**	0,29	0,13
	Stdev	1,55	0,69
	Median	0	0
	Max	31	12
Browsing 2	**Average**	1,9	0,76
	Stdev	4,74	3,56
	Median	0	0
	Max	53	73
Browsing 3	**Average**	0,84	0,62
	Stdev	2,14	2,11
	Median	0	0
	Max	27	50
Checkout	**Average**	2,34	1,32
	Stdev	5,19	4,96
	Median	0	0
	Max	58	88
Checkout-Registration	**Average**	0,22	0,09
	Stdev	0,79	0,51
	Median	0	0
	Max	8	9
Contact	**Average**	1,68	0,55
	Stdev	3,15	3,72
	Median	0	0
	Max	15	67
Faq	**Average**	0,09	0,04
	Stdev	0,41	0,28
	Median	0	0
	Max	5	8
MyAccount Page	**Average**	2,81	0,9
	Stdev	6,16	4,56
	Median	0	0
	Max	71	81
Newsletter Registration	**Average**	0,03	0,08
	Stdev	0,29	0,48
	Median	0	0
	Max	4	10
Other Page	**Average**	7,97	3,68
	Stdev	10,41	11,16
	Median	4	0
	Max	93	134
Price	**Average**	0,24	0,4
	Stdev	1,26	1,36
	Median	0	0
	Max	29	30
Product	**Average**	7,93	12,84
	Stdev	11,46	21,28
	Median	5	6
	Max	139	225
Registration	**Average**	0,28	0,08
	Stdev	1	0,54
	Median	0	0
	Max	8	12
salescycle_duration	**Average**	230,26	657,55
	Stdev	142,79	187,24
	Median	258,5	763,8
	Max	811,45	1018,62
total_pagehits	**Average**	28,29	22,27
	Stdev	31,89	38,76
	Median	18	8
	Max	282	294
opted_out	**Percentage opted out**	0,22	3,14

Tabelle 13: Statistiktabelle für die Datenanalyse
Quelle: Eigene Darstellung

Beim Betrachten der Statistiktabelle lässt sich feststellen, dass es keine Aktionen gibt, die deutlich für eine erhöhte Kaufwahrscheinlichkeit sprechen. Konvertierte Leads führen in der Tendenz allerdings mehr Aktionen aus. URLs der Kategorien "Basket", "Browsing 1-3",

"Checkout", "Checkout-Registration", "Contact", "Faq", "Other Page" und "Registration" werden dabei tendenziell häufiger von konvertierten Leads aufgerufen. URLs der Kategorien "Newsletter-Registration", "Price" und "Product", werden tendenziell häufiger von Kunden aufgerufen, die nicht konvertieren. Außerdem melden sich fast ausschließlich Kunden, die nicht konvertieren, aus dem Newsletter ab. Außerdem sinkt die Wahrscheinlichkeit, dass ein Kontakt zum Kunden wird, sofern dessen Erfassung länger zurückliegt.

Zusätzlich zur Statistiktabelle wurde eine Korrelationsanalyse der einzelnen Spalten aus dem Leadübersicht DF mit der Konversion zum Kunden durchgeführt. Die einzelnen Korrelationen sind Abbildung 18 zu entnehmen.

Abbildung 18: Korrelationsanalyse mit dem Lead-Status
Quelle: Eigene Darstellung

Die Ergebnisse aus der Korrelationsanalyse bestätigen die Schlüsse, welche aus der Statistiktabelle gezogen wurden. Der Code zur Datenanalyse ist in Anhang 5 zu entnehmen.

Entwicklung des Punktesystems

Zum Erstellen des Punktesystems werden folgende Schritte durchgeführt:

Erstellen der Scores

Für das Scoring im Business-to-Consumer (B2C)- beziehungsweise E-Commerce-Bereich ist zu beachten, dass im Vergleich zum B2B-Bereich in der Regel andere explizite Parameter wie Geschlecht, Alter und Wohnort relevant sind (vgl. Wuttke o. D.). Da diese Parameter

jedoch in der Datenbank des Onlineshops nicht zur Verfügung stehen, wird lediglich ein impliziter Score angelegt.

Festlegen der impliziten Bepunktung

Beim Festlegen der Bepunktung wird basierend auf der Datenanalyse zunächst entschieden, ob für eine Aktion Punkte hinzugefügt oder abgezogen werden. Im Anschluss werden auf einer Skala von eins bis zehn Punkte für die einzelnen Aktionen vergeben. Punkte, welche Indikatoren für ein unmittelbares Kaufinteresse oder kein Interesse an einem Kauf sind, wurden darüber hinaus mit dem Faktor 2,5 multipliziert. Als Indikatoren für ein aktuelles Kaufinteresse werden Websitebesuche vom Typ "Checkout" und "Checkout-Registration" gewählt. Diese gehen zwar aus der Datenanalyse nicht als eindeutige Kaufindikatoren hervor, sind allerdings der letzte Schritt im Kaufprozess, nachdem ein Kunde im Einkaufswagen auf "Kaufen" klickt. Somit deuten sie darauf hin, dass Kaufinteresse besteht und der Kontakt bereits kurz vor dem Kauf stand. Eine Aktion, die darauf hinweist, dass kein Kaufinteresse besteht, ist die Abmeldung vom Newsletter. Das gewählte Punktesystem ist in Tabelle 14 ersichtlich.

Pagehit Type	Stats	Converted	Unconverted	Radio (Converted / Unconverted)	Postive / Negative	Points	Indicator for immediate purchase interest / non purchase interest	Final Points
Basket	Average	1,67	0,78	2,14	+	10	No	10
Browsing 1	Average	0,29	0,13	2,23	+	1	No	1
Browsing 2	Average	1,9	0,76	2,50	+	1	No	1
Browsing 3	Average	0,84	0,62	1,35	+	1	No	1
Checkout	Average	2,34	1,32	1,77	+	10	Yes	25
Checkout-Registration	Average	0,22	0,09	2,44	+	10	Yes	25
Contact	Average	1,68	0,55	3,05	+	3	No	3
Faq	Average	0,09	0,04	2,25	+	2	No	2
Account	Average	2,81	0,9	3,12	+	4	No	4
Newsletter Registration	Average	0,03	0,08	0,38	+	3	No	3
Other Page	Average	7,97	3,68	2,17	+	1	No	1
Price	Average	0,24	0,4	0,60	+	5	No	5
Product	Average	7,93	12,84	0,62	+	4	No	4
Registration	Average	0,28	0,08	3,50	+	3	No	3
Optout	% opted out	0,22	3,14	0,07	-	-10	Yes	-25

Tabelle 14: Punktesystem im traditionellen Lead-Scoring-Praxisbeispiel
Quelle: Eigene Darstellung

Erstellen eines Unterdrückungssegments

Um sicherzustellen, dass keine irrelevanten Kontakte als MQLs eingestuft werden, wird ein Unterdrückungssegment gebildet.

Wahl des Verfallsmodells

Da aktuell noch nicht wissenschaftlich untersucht wurde, ob durch die Nutzung eines Verfallsmodells tatsächlich verbesserte Resultate erzielt werden, wird im Anwendungsbeispiel auf ein Verfallsmodell verzichtet. Da kein Verfallsmodell besteht, welches den Score auf einen negativen Bereich absenken kann, wird keine Kampagne zum Zurücksetzen negativer Scores benötigt.

Festlegen des Schwellenwertes

Mithilfe des Codes aus Anhang 6 wird rückwirkend für konvertierte Leads des Onlineshops der Score zum Zeitpunkt des ersten Kaufs und für nicht-konvertierte Leads deren aktueller Score errechnet. Zur Analyse der Schwellenwerte wurden zwei Histogramme erstellt (siehe Abbildung 19), welche die Schwellenwerte konvertierter- und nicht-konvertierter Leads vergleichen. Daraus geht hervor, dass konvertierte Leads zum Zeitpunkt der Übergabe tendenziell höhere Punktestände haben als Leads, die nicht konvertieren.

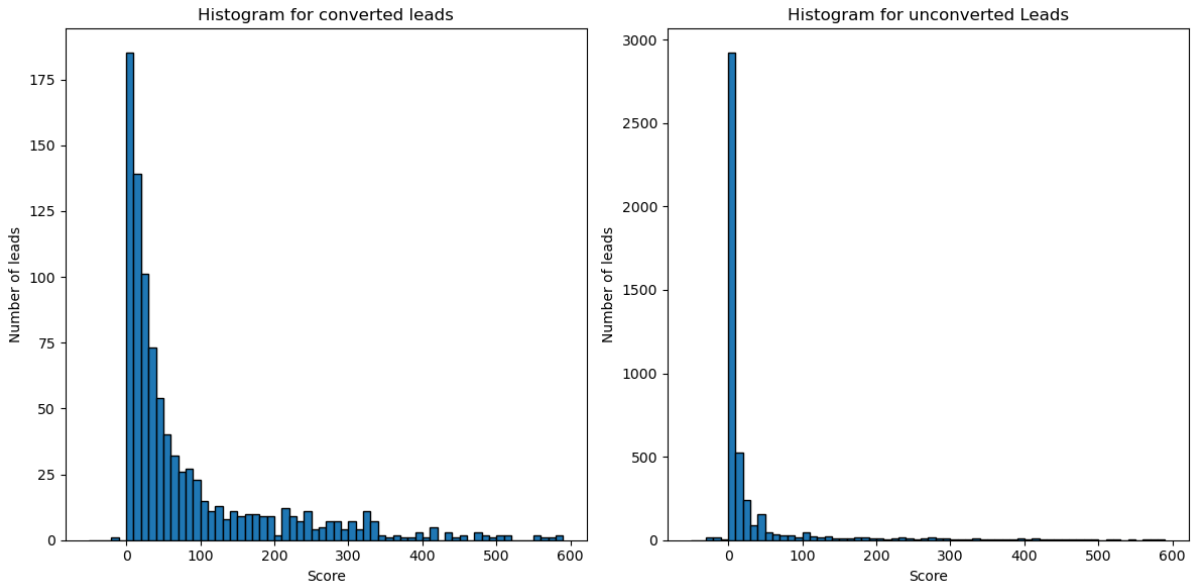

Abbildung 19: Vergleich der Schwellenwerte konvertierter und nicht-konvertierter Leads
Quelle: Eigene Darstellung

Außerdem wurde eine Tabelle zur Analyse der Perzentile erstellt (vgl. Tabelle 15). Aus dieser Tabelle geht hervor, dass bei einem Schwellenwert von 54 Punkten 50 Prozent der konvertierten Leads korrekt vorhergesagt werden und weniger als 25 Prozent der nicht-konvertierten Leads fälschlicherweise als konvertierende Leads vorhergesagt werden.

	Percentile	Converted_0	Converted_1
0	0.00	-23.0	1.0
1	0.05	4.0	5.0
2	0.10	8.0	9.0
3	0.15	8.0	13.0
4	0.20	8.0	17.0
5	0.25	8.0	18.0
6	0.30	9.0	24.0
7	0.35	12.0	31.0
8	0.40	15.0	36.0
9	0.45	16.0	48.0
10	0.50	18.0	54.0
11	0.55	21.0	69.0
12	0.60	24.0	86.0
13	0.65	32.0	103.0
14	0.70	40.0	123.0
15	0.75	52.0	156.0
16	0.80	74.0	207.0
17	0.85	130.0	273.0
18	0.90	207.0	350.0
19	0.95	410.0	574.0

Tabelle 15: Tabelle zur Analyse der Score-Perzentile
Quelle: Eigene Darstellung

Im Anschluss wurde der Schwellenwert schrittweise angepasst und ermittelt, bei welchem Wert die besten Ergebnisse erzielt werden. Hierbei wurde ein Schwellenwert von 52 Punkten ermittelt, bei dem der Status der Konvertierung mit einer Genauigkeit von 71,73 Prozent vorhergesagt wird. Die Ergebnisse bei einem Schwellenwert von 52 sind Abbildung 20 zu entnehmen.

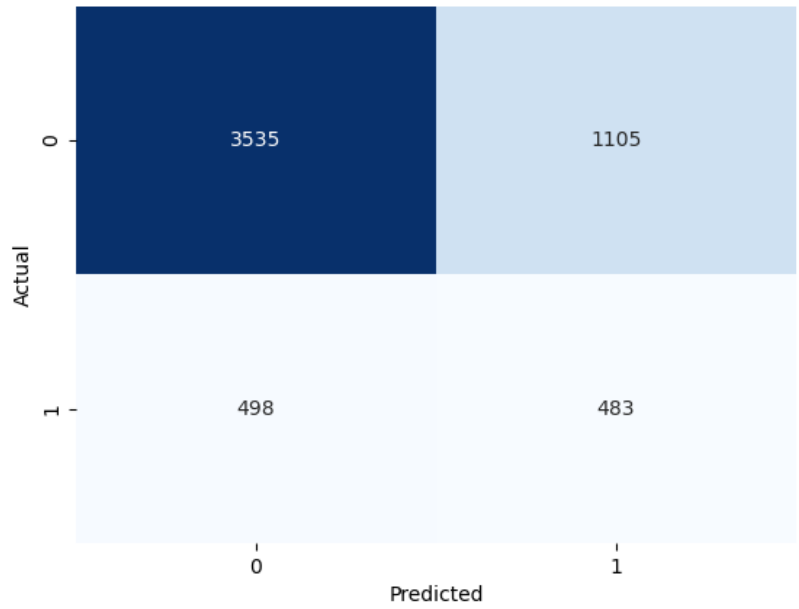

Abbildung 20: Konfusionsmatrix des vorläufigen traditionellen Lead-Scoring-Systems
Quelle: Eigene Darstellung

Anpassen des Lead-Scoring-Systems

Beim Anpassen des Systems wurde die Bepunktung mehrfach geändert und dazu der passende Schwellenwert ermittelt. Die besten Vorhersagen mit 76,55 Prozent Genauigkeit wurden erzielt, indem die Bepunktung mehr an die Ergebnisse der Datenanalyse angepasst wurde. Hierzu wurden die Punkte für den Besuch einer Seite der Kategorie "Price" auf zwei

Punkte abgesenkt und die Punkte für den Besuch einer Seite der Kategorie "Product" auf einen Punkt reduziert. Als Schwellenwert wurde 34 gewählt. Die Ergebnisse der Konfusionsmatrix des angepassten Systems sind Abbildung 21 zu entnehmen.

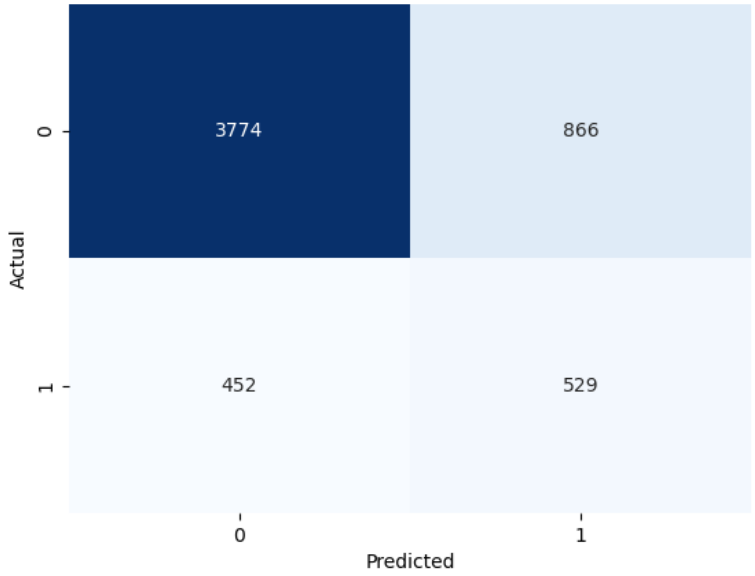

Abbildung 21: Konfusionsmatrix des finalen traditionellen Lead-Scoring-Systems
Quelle: Eigene Darstellung

Senden von Marketingmaßnahmen für MQLs

Da kein Vertriebsteam besteht, das MQLs weiterqualifiziert und kontaktiert, erfolgt nach dem Erreichen des Schwellenwertes keine Übergabe an das Vertriebsteam. Alternative Aktionen zur Übergabe an das Vertriebsteam für MQLs sind das Erstellen von MQL-Segmenten, das Senden von speziellen Angeboten, CTAs und Marketinginhalten für MQLs oder die Ansprache von MQLs über soziale Medien.

Auswerten der Lead Scores

Da kein Feedback von einem Vertriebsteam angefordert werden kann, weicht der Schritt der Lead-Auswertung von dem im Abschnitt 7.2 beschriebenen ab. Allerdings kann erfasst werden, welche MQLs zu Kunden konvertieren. Auch das Erstellen eines Recycling-Pfades ist denkbar. Auf diesem können Leads, welche innerhalb einer zuvor festgelegten Zeit nach dem Erhalt der Marketingmaßnahmen für MQLs nicht konvertieren, erneut in den Lead-Scoring-Prozess eingegliedert werden.

Messen des Erfolgs

Nachdem das System in Betrieb genommen wurde, wird der Erfolg mittels KPIs gemessen. Dazu werden mehrere KPIs festgelegt, zum Beispiel diejenigen, die in Kapitel 6 beschrieben werden, und regelmäßig überprüft. Wenn sich die KPIs verschlechtern, muss möglicherweise das Scoring-System aktualisiert werden.

Aktualisieren des Lead-Scoring-Systems

Mindestens alle drei Monate wird das Lead-Scoring-System aktualisiert. Dazu werden die KPIs geprüft, die Scores der Leads analysiert, neue Marketingmaterialien in das System integriert, die Bewertung durch Analyse der konvertierten MQLs angepasst und der Schwellenwert neu festgelegt.

8.2 Prädiktives Lead Scoring

Erstellen eines Service Level Agreements

Da kein Kontakt zum Marketing- und Vertriebsteam des Onlineshops besteht, kann auch für das prädiktive Lead-Scoring-System kein SLA erstellt werden.

Datengenerierung

Als Datenbasis werden die Daten aus der Datenbank des Onlineshops verwendet. Wie beim traditionellen Lead Scoring stehen auch beim prädiktiven Lead Scoring die Daten aller Websitebesuche inklusive Zeitstempel, das Datum von Newsletterabmeldungen, das Datum der Lead-Erfassung, das Datum der Konvertierung und das Datum der letzten Aktivität zur Verfügung. Außerdem kann aus der Segmentzugehörigkeit der Kontakte abgeleitet werden, ob es sich um konvertierte oder nicht-konvertierte Leads handelt.

Datenaufbereitung

Die Datenaufbereitung wird innerhalb von Python durchgeführt. Hierzu werden zunächst die SQL-Statements aus Anhang 3 und 4 verwendet, um die Daten konvertierter Leads und nicht-konvertierter Leads aus Mautic zu exportieren. Im Anschluss werden wie beim traditionellen System die Schritte der Datenauswahl, Datenbereinigung, Datenkonstruktion, Datenintegration und Datenformatierung durchgeführt. Daraus ergibt sich der Leadübersicht DF aus Tabelle 12. Zudem werden eine Skalierung und eine Aufteilung der Daten in Trainings- und Testdaten vorgenommen.

Modellierung

Nachdem die Daten aufbereitet sind, wird ein robustes Machine-Learning-Modell entwickelt. Hierzu wurden zunächst die Algorithmen Logistic Regression, Decision Tree, Random Forest und Support Vector Machine ausgewählt, bei welchen es sich um gängige Machine-Learning-Algorithmen im prädiktiven Lead Scoring handelt (vgl. Kapitel 4.1). Für die Modellentwicklung wird für die zuvor identifizierten Algorithmen eine Grid-Search durchgeführt. Das bedeutet, dass die Modelle mit verschiedenen Kombinationen der ermittelten Hyperparameter trainiert und getestet werden, um die besten Einstellungen zu finden. Zusätzlich wird die Genauigkeit der Modelle mit den jeweils besten Hyperparametern berechnet, die angibt, wie viel Prozent der Vorhersagen eines Modells korrekt sind. Dabei wurden folgende Genauigkeiten erreicht:

- Random Forest: 94,93 Prozent
- Decision Tree: 94,76 Prozent
- Support Vector Machine: 92,18 Prozent
- Logistic Regression: 88,0 Prozent

Folglich wurde das Random Forest Modell mit den aus der Grid-Search ermittelten Parametern gewählt. Die Ergebnisse des Random-Forest-Modells in Form einer Konfusionsmatrix, welche die Vorhersage mit dem tatsächlichen Status "konvertiert" (1) und "nicht konvertiert" (0) vergleicht, sind Abbildung 22 zu entnehmen. Die Konfusionsmatrizen zu den Algorithmen Logistic Regression, Decision Tree und Support Vector Machine sind den Anhängen 7, 8 und 9 zu entnehmen. Im Anschluss an die Grid-Search wird das beste Modell gespeichert, damit dieses später verwendet werden kann, um Vorhersagen zu treffen, ohne das Modell erneut trainieren zu müssen. Der Code zur Modellierung und zum Speichern des Modells ist im Anhang 10 zu finden.

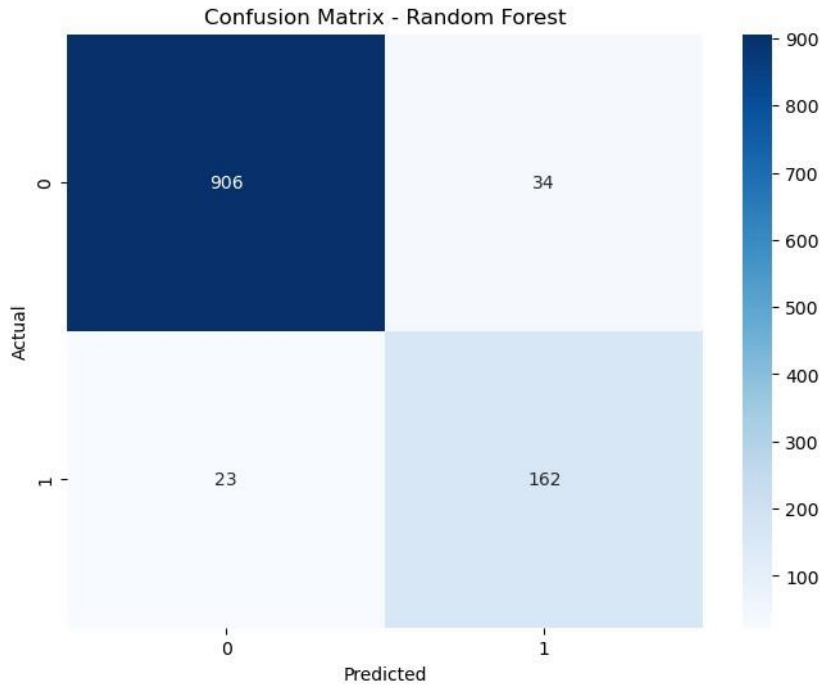

Abbildung 22: Konfusionsmatrix zum prädiktiven Lead Scoring mit dem Random-Forest-Algorithmus
Quelle: Eigene Darstellung

Modell-Evaluation

In dieser Evaluationsphase wurde nach Schwachstellen im Modell und in dessen Konstruktionsprozess gesucht. Hierbei wurden einige Anpassungen vorgenommen, welche bereits im Prozess der Datenaufbereitung berücksichtigt sind.

Modell-Implementierung

Die folgenden Schritte beschreiben die Anwendung des zuvor gespeicherten Modells auf Mautic:

Erstellen der Vorhersagen

Um die Vorhersagen zu erstellen, werden alle Leads, welche sich nicht im Unterdrückungssegment befinden, aus Mautic exportiert und aufbereitet. Im Anschluss wird das zuvor gespeicherte Machine-Learning-Modell geladen und damit Vorhersagen zu den einzelnen Leads getroffen. Die Vorhersagen werden gemeinsam mit der zugehörigen Lead-ID als CSV-Datei gespeichert. Der Code hierzu ist unter Anhang 11 zu finden.

Übertragen der Modellvorhersagen in Mautic

Nachdem die Vorhersagen generiert wurden, wird die CSV-Datei mit den Modellvorhersagen in Mautic importiert. Hierzu wird beim Import die Spalte lead_id der Mautic Contact ID zugeordnet und die Spalte mit den Vorhersagen einem zuvor für die Vorhersagen erstellen Datenbankfeld.

Senden von Marketingmaßnahmen für MQLs

Da kein Vertriebsteam vorhanden ist, um MQLs weiter zu qualifizieren und zu kontaktieren, erfolgt nach dem Überschreiten des Schwellenwerts keine Übergabe an das Vertriebsteam. Alternative Maßnahmen zur Übergabe von MQLs an das Vertriebsteam sind das Erstellen von MQL-Segmenten, das Senden von speziellen Angeboten, CTAs und Marketinginhalten für MQLs oder die Ansprache von MQLs über soziale Medien.

Auswerten der Leads

Da kein Feedback von einem Vertriebsteam angefordert werden kann, weicht der Schritt der Lead-Auswertung von dem im Abschnitt 7.2 beschriebenen ab. Allerdings kann erfasst werden, welche MQLs zu Kunden konvertieren. Auch das Erstellen eines Recycling-Pfades ist denkbar. Auf diesem können Leads, welche innerhalb einer zuvor festgelegten Zeit nach dem Erhalt der Marketingmaßnahmen für MQLs nicht konvertieren, erneut in den Lead-Scoring-Prozess eingegliedert werden.

Messung des Erfolgs

Nachdem das System in Betrieb genommen wurde, wird dessen Erfolg mittels KPIs gemessen. Dazu werden mehrere KPIs festgelegt, zum Beispiel diejenigen, die in Kapitel 6 beschrieben werden, und regelmäßig überprüft. Verschlechtern sich die KPIs, so muss das Machine-Learning-Modell möglicherweise aktualisiert werden.

Aktualisierung des Lead-Scoring-Systems

Mindestens alle drei Monate wird das Lead-Scoring-System aktualisiert. Hierzu werden die KPIs geprüft, die Ergebnisse näher analysiert und gegebenenfalls ein neues Machine-Learning-Modell antrainiert.

9 Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurde ein Lead-Scoring-Modell für einfache und erweiterte Anwendungsfälle am Beispiel der Software Mautic entwickelt. Hierzu wurde nach einer Literaturrecherche zunächst ein generisches Vorgehensmodell zur Entwicklung eines robusten traditionellen und prädiktiven Lead-Scoring-Systems erarbeitet. Im Anschluss wurde das Modell an die Software Mautic angepasst, wobei auch die Anwendungsfälle des produktbasierten Lead Scorings und des ABS berücksichtigt wurden. Abschließend wurden der traditionelle und der prädiktive Lead-Scoring-Ansatz und deren Ergebnisse anhand eines Praxisbeispiels untersucht. Die einzelnen Ansätze und Anwendungsfälle des Lead Scorings, deren praktische Umsetzung, die daraus resultierenden Ergebnisse, deren Interpretation, die identifizierten Optimierungspotenziale für die Software Mautic, die Limitationen der Forschung sowie Empfehlungen für weiterführende Forschungen werden im Folgenden zusammenfassend dargestellt.

Traditionelles Lead Scoring

Beim traditionellen Lead-Scoring-Ansatz erhält jeder Lead einen Score. Diesem werden je nach Eigenschaften und Verhalten des Leads Punkte hinzugefügt oder abgezogen. Überschreitet der Score eines Kontakts einen vordefinierten Schwellenwert, gilt er als qualifiziert und wird an das Vertriebsteam weitergeleitet oder erhält spezielle Marketingmaßnahmen für MQLs. Um ein datenbasiertes, traditionelles Lead-Scoring-System in Mautic aufzubauen, werden die Daten aus der Datenbank der Software exportiert und extern, zum Beispiel mit Python-Code, analysiert. Auf Basis der Datenanalyse werden von Experten eine implizite und eine explizite Scorecard erstellt. Letztere definiert die Punkte, die vergeben werden, wenn ein Lead vordefinierte Aktionen ausführt oder bestimmte Kriterien erfüllt. Leads, welche die impliziten und expliziten Schwellenwerte überschritten haben, werden anschließend an das zuständige Vertriebsteam übergeben oder mit Marketingmaterialien für MQLs versorgt. Abschließend wird Feedback zu den einzelnen MQLs generiert, um das System zu optimieren oder Kontakte, die noch nicht kaufbereit sind, erneut in den Scoring-Prozess einzubinden.

Prädiktives Lead Scoring

Neben dem traditionellen Lead-Scoring-Ansatz existiert auch der prädiktive Ansatz, welcher mithilfe von Algorithmen des maschinellen Lernens Leads prognostiziert, die eine hohe Kaufwahrscheinlichkeit aufweisen. Beim prädiktiven Scoring werden die Lead-Daten aus Mautic exportiert, extern mittels Python-Code aufbereitet und anschließend in Trainings- und Testdaten aufgeteilt. Daraufhin werden die Trainings- und Testdaten verwendet, um verschiedene Machine-Learning-Modelle zu trainieren und herauszufinden, welcher Algorithmus die besten Ergebnisse liefert. Das Modell, das die präzisesten Vorhersagen trifft,

wird dann gespeichert, um es für zukünftige Vorhersagen zu verwenden. Um Vorhersagen für neue Leads zu treffen, werden deren Daten zunächst aus Mautic importiert und aufbereitet. Anschließend wird das gespeicherte Modell auf die Daten angewandt. Die Vorhersagen des Modells werden dann in Mautic importiert. Leads, bei denen durch den Machine-Learning-Algorithmus eine Konvertierung zum Kunden vorhergesagt wird, werden nun an das Vertriebsteam weitergegeben oder mit Marketingmaterialien für MQLs versorgt. Auch hier wird Feedback gesammelt, um das Modell zu verbessern oder um Leads, die noch nicht bereit für einen Kauf sind, erneut in den Scoring-Prozess zu integrieren.

Es gilt zu beachten, dass sowohl für die im Rahmen dieser Arbeit vorgeschlagenen Implementierungen des traditionellen als auch des prädiktiven Ansatzes auf vorhandene Daten als Grundlage zurückgegriffen wird. Sollten diese Daten nicht verfügbar sein, besteht die Möglichkeit, solange bis ausreichend Daten für eine Datenanalyse zur Verfügung stehen, ein Modell zu entwickeln, das auf subjektiven Experteneinschätzungen basiert.

Produktbasiertes Lead Scoring

Das produktbasierte Lead Scoring ist ein Anwendungsfall des Lead Scorings, welcher darauf abzielt, das Interesse eines Leads an den einzelnen Produkten, Produktgruppen und Geschäftsbereichen eines Unternehmens zu bewerten und nicht das Interesse am Unternehmen als Ganzes. Dazu wird für jedes Produkt ein eigenes Scoring-System entwickelt. Um ein produktbasiertes Scoring-System in Mautic aufzubauen, kann sowohl dem traditionellen Ansatz als auch dem prädiktiven Ansatz gefolgt werden.

Account Based Scoring

Beim Lead-Scoring-Anwendungsfall des ABS werden Unternehmen anstelle einzelner Leads bewertet. Dazu wird im Zuge des traditionellen Ansatzes für jedes Unternehmen ein expliziter und ein impliziter Score gebildet. Der explizite Score setzt sich aus Unternehmensmerkmalen wie Größe, Branche oder technologischem Stand zusammen. Der implizite Score setzt sich aus der Summe der Aktionen zusammen, welche die einzelnen einem Unternehmen zugeordneten Leads durchführen. Da die Funktionalitäten des Account Based Marketing in Mautic derzeit noch nicht fortgeschritten genug sind, kann dort derzeit kein automatisiertes ABS betrieben werden. Dies ist sowohl bei der Anwendung des traditionellen als auch des prädiktiven Ansatzes im Rahmen des ABS der Fall.

Zusammenfassung und Interpretation der Ergebnisse aus dem Praxisbeispiel

Im traditionellen Lead Scoring konnte mit einer Genauigkeit von 76,55 Prozent vorhergesagt werden, welche Leads zu Kunden werden. Im Gegensatz dazu wurden beim prädiktiven Lead Scoring vier verschiedene Machine-Learning-Algorithmen untersucht, wobei der Random-Forest-Algorithmus mit einer Genauigkeit von 94,93 Prozent die besten Ergebnisse

erzielte. Diese Resultate unterstützen die Erkenntnisse anderer Forschungsarbeiten, die das prädiktive Lead Scoring als effektivere Alternative zum traditionellen Ansatz identifizieren. Dadurch, dass die Leistungsfähigkeit der beiden Ansätze am gleichen Anwendungsbeispiel getestet wurde, kann diese These bekräftigt werden.

Diese Ergebnisse sind darauf zurückzuführen, dass das prädiktive Lead Scoring einige der Schwächen des traditionellen Lead Scorings abdeckt. So können beispielsweise nicht-lineare Effekte in prädiktiven Systemen berücksichtigt werden. Außerdem wird die Subjektivität aus dem Bewertungsprozess genommen, da sich das System vollständig auf Daten stützt. Ein weiterer Vorteil des prädiktiven Scorings ist die Reduktion der Abhängigkeit von der Menge an verhaltensbasierten Daten, weshalb Kontakte, die von Anfang an ein Kaufinteresse aufweisen, schneller identifiziert werden.

Die Erkenntnis, dass der Random-Forest-Algorithmus die besten Ergebnisse im prädiktiven Lead Scoring erzielt, findet Bestätigung in mehreren der untersuchten Forschungsarbeiten. Es bleibt jedoch unsicher, ob eine erhöhte Genauigkeit der Modelle zwangsläufig zu verbesserten Ergebnissen führt. Eine gesteigerte Genauigkeit führt zwar dazu, dass ein höherer Prozentsatz an qualifizierten Leads an das Vertriebsteam weitergeleitet wird, was die Conversion Rate von MQLs zu Kunden steigert. Dennoch werden beim Lead Scoring auch andere KPIs wie die Auswirkungen auf den Umsatz oder den Gewinn gemessen. Es ist ungewiss, ob eine höhere Genauigkeit automatisch zu Verbesserungen dieser KPIs führt. Beispielsweise könnte bei einer Genauigkeit von 95 Prozent der Fokus vorwiegend auf MQLs liegen, die ohnehin kaufen würden. Dies hätte zur Folge, dass nur wenige Leads mit zwar vorhandenem Kaufpotenzial, aber geringerer Konversionswahrscheinlichkeit übergeben werden. Eine Lösung für dieses mögliche Problem besteht darin, mehrere Gruppen anhand der Leadqualität zu definieren. Beispielsweise können Leads zum Trainieren eines maschinellen Lernmodells in folgende Gruppen eingeteilt werden:

- Leads, die nicht übergeben wurden
- Leads, die übergeben wurden, aber nicht gekauft haben
- Leads, die übergeben wurden und konvertiert sind,

Dadurch ordnet das Modell die Leads nach dem Antrainieren in die drei unterschiedlichen Kategorien ein, die entsprechend ihrer Priorität bearbeitet werden können. Alternativ ist auch eine Bewertung der Leads durch das Vertriebsteam auf einer Skala von eins bis zehn denkbar, um das Modell zu entwickeln. Ein analoges Vorgehen kann im traditionellen Lead Scoring durch die Nutzung einer Lead-Scoring-Matrix, wie in Abbildung 2 dargestellt, umgesetzt werden. Hierbei werden Leads basierend auf der Höhe des impliziten und

expliziten Scores in verschiedene Kategorien eingeteilt, um sie dann schrittweise abzuarbeiten.

Des Weiteren ist zu berücksichtigen, dass sowohl für die im Rahmen dieser Arbeit vorgeschlagenen Implementierungen des traditionellen als auch des prädiktiven Ansatzes auf vorhandene Daten als Grundlage zurückgegriffen wird. Sollten diese Daten nicht verfügbar sein, besteht die Möglichkeit, vorläufig ein Modell zu entwickeln, das auf subjektiven Experteneinschätzungen basiert und dieses anzupassen, sobald ausreichend Daten für eine effektive Datenanalyse zur Verfügung stehen.

Empfehlungen für die Weiterentwicklung der Software Mautic

Es wurde festgestellt, dass es beim Verwenden mehrerer Scores, beispielsweise beim produktbasierten Scoring oder wenn mit einem impliziten und expliziten Score gearbeitet wird, zu Problemen kommen kann. Dies liegt daran, dass alle Point Groups, die neben dem Hauptscore erstellt werden, nur über eine provisorische Lösung in Form einer dedizierten Kampagne (siehe Abbildung 11) zurückgesetzt werden können, was die Komplexität des Scoring-Verfahrens erhöht. Außerdem gibt es keine Möglichkeit, den Wert einer Point Group durch den Import einer CSV-Datei festzulegen. Dies wäre insbesondere dann sinnvoll, wenn das Lead-Scoring-System mit bereits bestehenden Kunden implementiert wird und deren Scores rückwirkend ermittelt und in Mautic importiert werden sollen. Folglich gilt es, die Software so weiterzuentwickeln, dass auf Point Groups die gleichen Funktionen angewandt werden können wie auf den in Mautic vordefinierten Standard-Score eines Leads.

In Mautic muss außerdem ein Teil des Bepunktungssystems durch Kampagnen abgebildet werden, anstatt die Logik ausschließlich unter dem dafür vorgesehenen Menüpunkt "Manage Actions" zu implementieren. Dies gilt zum einen für explizite Parameter, da derzeit nur auf Basis von Aktionen und nicht auf Basis der Eigenschaften eines Leads Punkte vergeben werden können. Darüber hinaus sind auch einige Aktionen nicht im Menüpunkt "Manage Actions" enthalten, weshalb ein Teil der impliziten Bewertung über Kampagnen erfolgen muss. Ein Beispiel hierfür ist, wenn sich ein Kontakt aus einem Newsletter abmeldet. Daher ist eine weitere Empfehlung zur Optimierung der Software, sämtliche Aktionen und Eigenschaften, die den Score beeinflussen können, unter "Manage Actions" zu integrieren.

Im vorgeschlagenen Modell wird mit Unterdrückungssegmenten und mehreren Scores gearbeitet. Da jedoch unter dem Menüpunkt "Manage Triggers", der für das Auslösen von Aktionen beim Überschreiten von Score-Schwellenwerten vorgesehen ist, nur Aktionen auf der Grundlage eines einzelnen Scores gestartet werden können, ist es nicht möglich, mehrere Scores und Unterdrückungssegmente in der Logik zu implementieren. Alternativ müssen Leads Segmenten mit Filtern hinzugefügt werden, auf deren Grundlage dann

Kampagnen gestartet werden können. Daher wäre es sinnvoll, den Menüpunkt "Manage Triggers" so zu erweitern, dass auch kombinierte Point Triggers verwendet werden können. Diese könnten Schwellenwerte mehrerer Scores in Kombination mit weiteren Bedingungen wie beispielsweise Segmentmitgliedschaften berücksichtigen.

Des Weiteren ist es in Mautic zwar möglich, Unternehmen auf Account-Ebene zu bewerten, allerdings besteht keine Möglichkeit, darauf basierend automatisierte Aktionen durchzuführen. Um den Aufbau eines accountbasierten Scoring-Systems zu ermöglichen, wird im Rahmen dieser Arbeit empfohlen, die Software so weiterzuentwickeln, dass das accountbasierte Marketing und Scoring besser unterstützt wird.

Im Verlauf dieser Arbeit wurde sowohl im traditionellen als auch im prädiktiven Lead Scoring auf Python-Code zurückgegriffen. Daher könnte die Integration von Python in Mautic den Nutzern einen zusätzlichen Mehrwert bieten, indem sie prädiktives Lead Scoring und andere komplexe Analysen ohne manuellen Aufwand automatisieren und in Echtzeit durchführen könnten. Eine benutzerfreundlichere Alternative zur Einbindung von Python besteht darin, zusätzliche Datenanalyse- und Machine-Learning-Funktionen direkt in die Software zu integrieren. Dadurch kann das im Rahmen dieser Arbeit präsentierte Modell genutzt werden, ohne dass Programmierkenntnisse erforderlich sind.

Limitationen der Forschung

Im Rahmen der untersuchten Praxisbeispiele wurde die Genauigkeit der Vorhersagen der unterschiedlichen Ansätze ermittelt. Allerdings war es im Rahmen dieser Arbeit nicht möglich, das vorgeschlagene Modell im Unternehmen zu implementieren und die Veränderung der einzelnen KPIs zu messen. Daher wäre es von wissenschaftlicher Bedeutung, sowohl ein traditionelles als auch ein prädiktives Lead-Scoring-System in einem Unternehmen zu implementieren und deren Ergebnisse über einen längeren Zeitraum zu untersuchen und zu vergleichen.

Des Weiteren wurde das ABS im Praxisbeispiel nicht getestet, da dieser Anwendungsfall derzeit nicht in Mautic abgebildet werden kann und keine Unternehmensdaten zur Verfügung standen, da der Onlineshop aus dem Praxisbeispiel im B2C-Bereich tätig ist.

Auch der Test des produktbasierten Anwendungsfalls des Lead Scorings an einem Praxisbeispiel entfiel. Da im Rahmen des produktbasierten Scorings jedoch mehrere Lead-Scoring-Systeme gleichzeitig aufgebaut werden, ist davon auszugehen, dass diese je nach Wahl des Ansatzes ähnliche Ergebnisse wie der traditionelle und der prädiktive Ansatz erzielen.

Empfehlungen für weiterführende Forschungen

Im Rahmen der Untersuchung wurde festgestellt, dass es sich bei prädiktiven Lead-Scoring-Systemen um Modelle handelt, deren Vorgehen für die Anwender nicht nachvollziehbar ist. Obwohl mittlerweile Wege erforscht wurden, um das Vorgehen innerhalb der Modelle nachvollziehbarer zu machen, ist das genaue Vorgehen des Modells nach wie vor nicht transparent. Dieses Problem könnte durch weitere Forschung behoben werden, indem ein hybrider Ansatz aus traditionellem und prädiktivem Lead Scoring entwickelt wird, der mittels maschinellen Lernens das Punktesystem und den Schwellenwert für ein traditionelles Lead-Scoring-System entwickelt. Dadurch wäre das Modell vollständig datengetrieben und gleichzeitig nachvollziehbar. Zudem könnten manuelle Prozessschritte wie die Anpassung von Score und Schwellwert automatisiert werden, was Zeit bei der Aktualisierung des Systems spart.

Im Rahmen der Forschungsarbeit wurden außerdem mehrere Verfallsmodelle identifiziert, die dafür sorgen, dass der Score von Leads mit längerer Inaktivität herabgesetzt wird. Auf diese Weise haben nur Leads mit aktuellem Interesse einen hohen Score. Obwohl Verfallsmodelle von verschiedenen Anbietern von MAS und CRM-Systemen empfohlen werden, gibt es keine wissenschaftlichen Untersuchungen darüber, ob sie tatsächlich zu besseren Ergebnissen führen. Die Fragen, inwieweit sich ein Verfallsmodell auf den Erfolg des Lead Scorings auswirkt, wie das richtige Verfallsmodell ausgewählt und wie dieses erfolgreich konfiguriert wird, bieten daher Stoff für eine separate wissenschaftliche Arbeit.

Darüber hinaus wurden in Kapitel 5 mit dem RFM-Modell und Empfehlungssystemen zwei alternative beziehungsweise ergänzende Methoden zum Lead Scoring identifiziert. Eine Forschungsarbeit, welche die Anwendung der alternativen Methoden mit dem Lead Scoring vergleicht und Rückschlüsse darüber zieht, in welchen Fällen welche Methode besser geeignet ist und wann sich die Anwendung eines hybriden Ansatzes empfiehlt, wäre daher ebenfalls von wissenschaftlicher Bedeutung.

Literaturverzeichnis

Abbasi, Ahmed; Sarker, Suprateek; Chiang, Roger H.L. (2016): Big Data Research in Information Systems: Toward an Inclusive Research Agenda, in: *Journal of the Association for Information Systems*, Jg. 17, Nr. 2, S. 1–32.

ActiveCampaign (o. D.): *Lead Scoring Best Practices (The Only Framework You Need to Get Started)*, [online] https://www.activecampaign.com/learn/guides/lead-scoring-best-practices-the-only-framework-you-need-to-get-started [20.11.2023].

Adobe (2019): *DER ULTIMATIVE LEITFADEN FÜR LEAD SCORING.*, [online] https://business.adobe.com/de/resources/guides/lead-scoring.html [20.11.2023].

Apple (o. D.): *Erlauben von Cookies in Safari auf dem Mac*, [online] https://support.apple.com/de-lu/guide/safari/ibrw850f6c51/mac [20.11.2023].

Auerochs, Jessica (2021): *Lead Scoring - Definition und Umsetzung in der B2B Praxis*, [online] https://www.marconomy.de/lead-scoring-definition-und-umsetzung-in-der-b2b-praxis-a-1eac3dcecab184cf86a6d9104a2147f3/ [20.11.2023].

Bagshaw, Anthony (2015): What is marketing automation?, in: *J Direct Data Digit Mark Pract*, Jg. 17, Nr. 2, S. 84–85.

Biegel, Bruce (2009): The current view and outlook for the future of marketing automation, in: *J Direct Data Digit Mark Pract*, Jg. 10, Nr. 3, S. 201–213.

Bohanec, Marko; Kljajić Borštnar, Mirjana; Robnik-Šikonja, Marko (2015): Integration of machine learning insights into organizational learning. A case of B2B sales forecasting, in: *Proceedings of 28th Bled eConference,* Bled, S. 338–352.

Bohanec, Marko; Kljajić Borštnar, Mirjana; Robnik-Šikonja, Marko (2017): Explaining machine learning models in sales predictions, in: *Expert Systems with Applications*, Jg. 71, S. 416–428.

Braun, Simone (2021): Valide Kundendaten – Das Fundament für Omni-Channel Marketing, in: Uwe Hannig (Hrsg.), *Marketing und Sales Automation*, Wiesbaden: Springer, S. 159–175.

Brevo (2023): *What is a Lead Scoring Model and How to Make One*, [online] https://www.brevo.com/blog/lead-scoring-model/ [06.12.2023].

Buckinx, Wouter; van den Poel, Dirk (2005): Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting, in: *European Journal of Operational Research*, Jg. 164, Nr. 1, S. 252–268.

Cleff, Thomas (2019): *Applied Statistics and Multivariate Data Analysis for Business and Economics. A Modern Approach Using SPSS, Stata, and Excel*, Cham: Springer International Publishing.

ConstantContact (2023): *Understanding Lead Score Decay*, [online] https://knowledgebase.constantcontact.com/lead-gen-crm/articles/KnowledgeBase/50268-Understanding-Lead-Score-Decay?lang=en_US [16.02.2024].

D'Haen, J.; van den Poel, D.; Thorleuchter, D.; Benoit, D. F. (2016): Integrating expert knowledge and multilingual web crawling data in a lead qualification system, in: *Decision Support Systems*, Jg. 82, S. 69–78.

D'Haen, Jeroen; van den Poel, Dirk (2013): Model-supported business-to-business prospect prediction based on an iterative customer acquisition framework, in: *Industrial Marketing Management*, Jg. 42, Nr. 4, S. 544–551.

Day, Daniel G.; Wei Shi, Savannah (2020): Automated and Scalable: Account-Based B2B Marketing for Startup Companies, in: *JBTP*, Jg. 8, Nr. 2, S. 16-23.

DMEXCO (o. D.): *Marketing Automation*, [online] https://dmexco.com/de/topic/marketing-automation/ [09.03.2024].

Duncan, Brendan; Eklan, Charles (2015): Probabilistic Modeling of a Sales Funnel to Prioritize Leads, in: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York: Association for Computing Machinery, S. 1751–1758.

Elkan, Charles (2013): *Predictive analytics and data mining*, University of California, San Diego.

Encharge.io (2021): *10 Must-Know Lead Scoring Best Practices to Win Your Leads in 2022*, [online] https://encharge.io/lead-scoring-best-practices/ [20.11.2023].

Espadinha-Cruz, P.; Fernandes, A.; Grilo, A. (2021): Lead management optimization using data mining: A case in the telecommunications sector, in: *Computers & Industrial Engineering*, Jg. 154, S. 1–14.

faraday.ai (o. D.): *Predictive lead scoring: Best practices for B2C lead scoring using machine learning*, [online] https://faraday.ai/blog/predictive-lead-scoring-b2c [20.11.2023].

Flocke, Louisa; Holland, Heinrich (2014): Die Customer Journey Analyse im Online Marketing, in: *Dialogmarketing Perspektiven 2013/2014*, Wiesbaden: Springer Fachmedien, S. 213–242.

Ghorbel, Amine (2023): *Everything you Need to Know About Lead Scoring in 2023*, [online] https://lagrowthmachine.com/lead-scoring/ [06.12.2023].

Gokhale, Prasad; Joshi, Pratima (2018): A Binary Classification Approach to Lead Identification and Qualification, in: A. V. Deshpande, Aynur Passi U.K., Dharm S.M. Nayak und Bharat P.S. Pathan (Hrsg.), *Smart Trends in Information Technology and Computer Communications*, Singapur: Springer, S. 279–291.

Gooding, Resa (2022): *Empowering Marketing and Sales with HubSpot. Take your business to a new level with HubSpot's inbound marketing, SEO, analytics, and sales tools*, 1. Aufl., Birmingham: Packt Publishing Limited.

Google (o. D.): *Cookies in Chrome löschen, zulassen und verwalten*, [online] https://support.google.com/chrome/answer/95647 [20.11.2023].

Gradow, Lisa; Greiner, Ramona (2021): *Quick Guide Consent-Management*, Wiesbaden: Springer.

Griebsch, Laura (2021): *Marketing Automation im B2B – Grundlagen und Erklärungen und 8 Tipps für den Einsatz*, [online] https://www.marconomy.de/marketing-automation-im-b2b-grundlagen-und-erklaerungen-und-8-tipps-fuer-den-einsatz-a-1001337/ [30.01.2024].

Hannig, Uwe (2020): Marketing-Automaton - automatisch mehr Markterfolg, in: Marcus Stumpf (Hrsg.), *Die 10 wichtigsten Zukunftsthemen im Marketing,* 2. Aufl., Freiburg: Haufe-Lexware GmbH & Co. KG, S. 207–230.

Hannig, Uwe (2021): Lead Management Automation vereint Marketing und Vertrieb, in: Uwe Hannig (Hrsg.), *Marketing und Sales Automation*, Wiesbaden: Springer, S. 243–257.

Heinzelbecker, Klaus (2021a): Account-based Marketing mit CRM und Marketing Automation, in: Uwe Hannig (Hrsg.), *Marketing und Sales Automation*, Wiesbaden: Springer, S. 387-410.

Heinzelbecker, Klaus (2021b): CRM, CXM und Marketing Automation, in: Uwe Hannig (Hrsg.), *Marketing und Sales Automation*, Wiesbaden: Springer, S. 135–147.

Hu, Hang; Peng, Peng; Wang, Gang (2019): *Characterizing Pixel Tracking through the Lens of Disposable Email Services*, *2019 IEEE Symposium on Security and Privacy (SP)*, San Francisco: IEEE, S. 365–379.

Hu, Jianfeng; Zhang, Bo (2012): *Product Recommendation System*, Stanford University.

Von der Hude, Marlis (2020): *Predictive Analytics und Data Mining*, Wiesbaden: Springer Fachmedien Wiesbaden.

Hufford, Brendan (2021): *Create an Effective Lead Scoring System in 7 Steps*, [online] https://www.activecampaign.com/blog/create-effective-lead-scoring-rules [06.12.2023].

InvestGlass (2023): *Die 4 besten Lead-Scoring-Modelle im Jahr 2023*, [online] https://www.investglass.com/de/the-4-best-lead-scoring-models-in-2023-examples/ [25.01.2024].

Jadli, Aissam; Hamim, Mohammed; Hain, Mustapha; Hasbaoui, Anouar (2022): TOWARD A SMART LEAD SCORING SYSTEM USING MACHINE LEARNING, in: *INDJCSE*, Jg. 13, Nr. 2, S. 433–443.

Jo, Taeho (2021): *Machine Learning Foundations. Supervised, Unsupervised, and Advanced Learning*, 2. Aufl., Cham: Springer International Publishing.

Jörvinen, Joel; Taiminen, Heini (2016): Harnessing marketing automation for B2B content marketing, in: *Industrial Marketing Management*, Jg. 54, S. 164–175.

Kazemi, Abolfazl; Babaei, Mohammad Esmaeil (2011): Modelling Customer Attraction Prediction in Customer Relation Management using Decision Tree: A Data Mining Approach, in: *Journal of Optimization in Industrial Engineering*, Jg. 4, Nr. 9, S. 37–45.

Kim, YongSeog; Street, W.Nick (2004): An intelligent system for customer targeting: a data mining approach, in: *Decision Support Systems*, Jg. 37, Nr. 2, S. 215–228.

Koerner, Alexander (2021): Marketing und Sales Automation, in: Uwe Hannig (Hrsg.), *Marketing und Sales Automation*, Wiesbaden: Springer, S. 61–78.

Kuhn, Max; Johnson, Kjell (2013): *Applied Predictive Modeling*, New York: Springer New York.

Kumar, V.; Reinartz, Werner (2018): *Customer Relationship Management*, Berlin, Heidelberg: Springer Berlin Heidelberg.

Lattice (2014): *The Evolution From Traditional to Predictive Lead Scoring: A how-to guide for considering predictive scoring*, [online] https://www.demandgenreport.com/industry-resources/ebooks/2892-the-evolution-from-traditional-to-predictive-lead-scoring [20.11.2023].

Lontzek, Nicole (2022): *Lead Scoring - Das Punktesystem für echte Marketingspezialisten*, [online] https://www.marconomy.de/lead-scoring-das-punktesystem-fuer-echte-marketingspezialisten-a-1087780/ [20.11.2023].

Martinez-Plumed, Fernando; Contreras-Ochando, Lidia; Ferri, Cesar; Hernandez-Orallo, Jose; Kull, Meelis; Lachiche, Nicolas; Ramirez-Quintana, Maria Jose; Flach, Peter (2021): CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories, in: *IEEE Trans. Knowl. Data Eng.*, Jg. 33, Nr. 8, S. 3048–3061.

mautic.org (o. D.): *About Mautic*, [online] https://github.com/mautic/mautic [08.12.2023].

Michiels, Ian (2008): *Lead Prioritization and Scoring. The Path to Higher Conversion*, [online] https://silo.tips/download/lead-prioritization-and-scoring [20.11.2023].

MM-Redaktion (2021): *Das nervt SEO-Agenturen: Apple schränkt Cookie-Tracking ein*, [online] https://magazinmedien.de/cookie-tracking_eingeschraenkt/ [20.11.2023].

Monat, Jamie P. (2011): Industrial sales lead conversion modeling, in: *Marketing Intelligence & Planning*, Jg. 29, Nr. 2, S. 178–194.

Mozilla (o. D.): *Cookies blockieren*, [online] https://support.mozilla.org/de/kb/Cookies-blockieren [20.11.2023].

Naveen, Kumar G.; Hariharanath, K. (2021): Designing a Lead Score Model for Digital Marketing Firms in Education Vertical in India, in: *Indian Journal of Science And Technology*, Jg. 14, Nr. 16, S. 1302–1309.

Nygard, Robert; Mezei, Jozsef (2020): Automating Lead Scoring with Machine Learning: An Experimental Study, in: *Proceedings of the 53rd Hawaii International Conference on System Sciences,* Honolulu, S. 1439–1448.

Oracle (o. D.): *Lead scoring*, [online] https://docs.oracle.com/en/cloud/saas/marketing/eloqua-user/Help/LeadScoring/LeadScoring.htm [25.01.2024].

Patel, Neil (o. D.): *Wie man Lead Scoring richtig nutzt, um mehr Einnahmen zu erzielen*, [online] https://neilpatel.com/de/blog/lead-scoring-richtig/ [05.12.2023].

Philipp, Martin (2021): Vom E-Mail-Marketing zum Lead Management, in: Uwe Hannig (Hrsg.), *Marketing und Sales Automation*, Wiesbaden: Springer, S. 193–213.

Rahimi, Omid (2020): *Mit 5 Tipps zum effektiven B2B-Lead-Scoring*, [online] https://www.marconomy.de/mit-5-tipps-zum-effektiven-b2b-lead-scoring-a-d2678f31de7c5524c335634a2e7e7df8/ [20.11.2023].

Saha, Swapan K.; Aman, Ashraful; Hossain, Md. shawkat; Islam, Aminul; Rodela, Ripa S. (2014): A Comparative Study On B2B Vs. B2C Based On Asia Pacific Region, in: *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, Jg. 3, Nr. 9, S. 294–298.

Salesforce (o. D.): *Scoring Categories*, [online] https://help.salesforce.com/s/articleView?id=sf.pardot_leadqual_scoring_categories.htm&type=5 [25.01.2024].

Schoepf, Alex (2021): Best Practices für Marketing-Automation-Einsteiger, in: *Wirtsch Inform Manag*, Jg. 13, Nr. 4, S. 280–289.

Schüller, Anne M.; Schuster, Norbert (2022): *Marketing-Automation. Neukundengewinnung, Up-Selling, Cross-Selling, Bestandskunden- Management: Mehr Umsatz mit der Wasserlochstrategie®*, 2. Aufl., Freiburg: Haufe Group.

Schuster, Norbert (2021): Marketing Automation ändert den Vertrieb, in: Uwe Hannig (Hrsg.), *Marketing und Sales Automation*, Wiesbaden: Springer, S. 105–119.

Schuster, Norbert (2022): *Digitalisierung in Marketing und Vertrieb. Richtige Strategien entwickeln und Potentiale der Digitalisierung für mehr Umsatz nutzen*, 2. Aufl., Freiburg, München, Stuttgart: Haufe Group.

Sereda, Evgeni (2021): *So gehst du mit URL-Parametern um und organisierst sie SEO-freundlich*, [online] https://www.semrush.com/blog/de/url-parameter/ [17.02.2024].

Sharma, Jatin; Sharma, Kartikay; Garg, Kaustubh; Sharma, Avinash Kumar (2021): Product Recommendation System a Comprehensive Review, in: *IOP Conf. Ser.: Mater. Sci. Eng.*, Jg. 1022, Nr. 1, S. 1–8.

Shearer, Colin (2000): The CRISP-DM Model: The New Blueprint for Data Mining, in: *Journal of Data Warehousing*, Jg. 5, Nr. 4, S. 13–22.

Simmoleit, Rainer (2023): *Künstliche Intelligenz im Lead Scoring*, [online] https://www.marconomy.de/kuenstliche-intelligenz-im-lead-scoring-a-5e21a051bb63bff701d8c39e72a2fef3/ [20.11.2023].

Swani, Lakshay; Tyagi, Prakita (2017): Predictive Modelling Analytics through Data Mining, in: *International Research Journal of Engineering and Technology*, Jg. 4, Nr. 9, S. 5–11.

Teiu, Codrin (2021): Marketing Automation Systems as Part of the Management Information Systems Evolution, in: *Organizations and Performance in a Complex World*, Cham: Springer, S. 325–333.

Todor, Raluca D. (2016): Marketing Automation, in: *Bulletin of the Transilvania University of Braşov*, Jg. 9, Nr. 2, S. 87–94.

Uhlemann, Ingrid Andrea (2015): *Einführung in die Statistik für Kommunikationswissenschaftler*, Wiesbaden: Springer Fachmedien Wiesbaden.

Verma, Rakesh; Koul, Saroj; Pai, Sushanth S. (2016): Identifying profitable clientele using the analytical hierarchy process, in: *International Journal of Business and Systems Research*, Jg. 10, 2-4, S. 220–237.

Wu, Migao; Andreev, Pavel; Benyoucef, Morad (2023): The state of lead scoring models and their impact on sales performance, in: *Information technology & management*, S. 1–30.

Wu, Ming-Wei; Lin, Ying-Dar (2001): Open source software development: an overview, in: *Computer*, Jg. 34, Nr. 6, S. 33–38.

Wuttke, Laurenz (o. D.): *Lead Scoring: Definition, Prozess und Unterschiede zwischen B2B vs. B2C*, [online] https://datasolut.com/lead-scoring [20.11.2023].

Zumstein, Darius; Gasser, Marc; Thüring, Urs; Völk, Klaus; Wicki, Alexander; Oswald, Carmen; Merdzanovic, Adis; Hannich, Frank (2023): *Marketing Automation Report 2023*, ZHAW Zürcher Hochschule für Angewandte Wissenschaften, Winterthur.

Anhang

Anhang 1: Kampagne zum Generieren von Vertriebsfeedback

Alternativ zur Implementierung der Logik zur Einholung von Vertriebsfeedback im CRM-System, kann hierzu auch eine Kampagne in Mautic verwendet werden (siehe Abbildung 23).

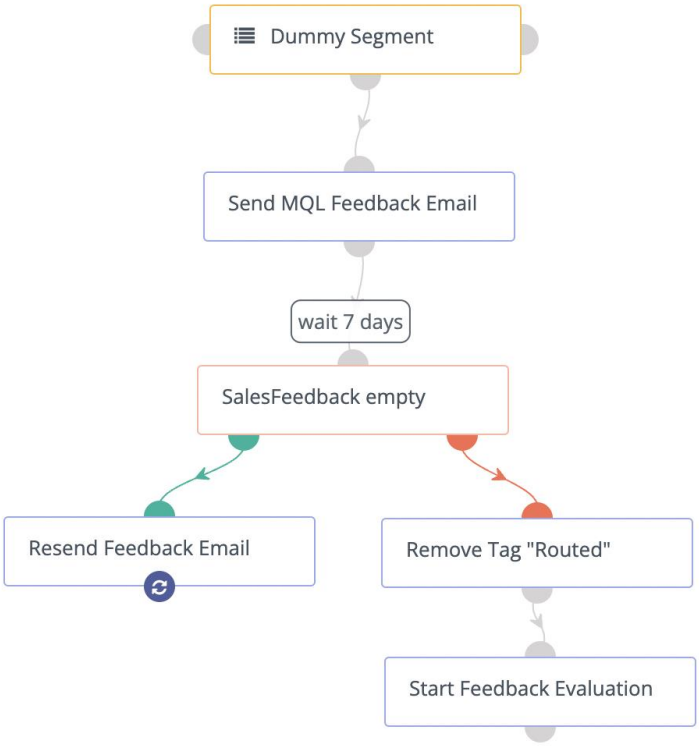

Abbildung 23: Kampagne zum Generieren von Vertriebsfeedback
Quelle: Eigene Darstellung

Zu Beginn dieser Kampagne wird eine E-Mail an das zuständige Vertriebsteam gesendet, in der um Feedback zur Qualität eines übergebenen Leads gebeten wird (siehe Abbildung 24). Innerhalb dieser E-Mail können die Vertriebsmitarbeiter auswählen, ob es sich bei dem Lead um einen SAL handelt, ob der Lead noch nicht bereit für den Vertriebsprozess ist und recycelt werden soll oder ob der Lead abgelehnt werden soll. Klickt ein Vertriebsmitarbeiter in der E-Mail auf einen der drei Feedback-Buttons, so wird das Feedback automatisch in der Datenbank von Mautic gespeichert.

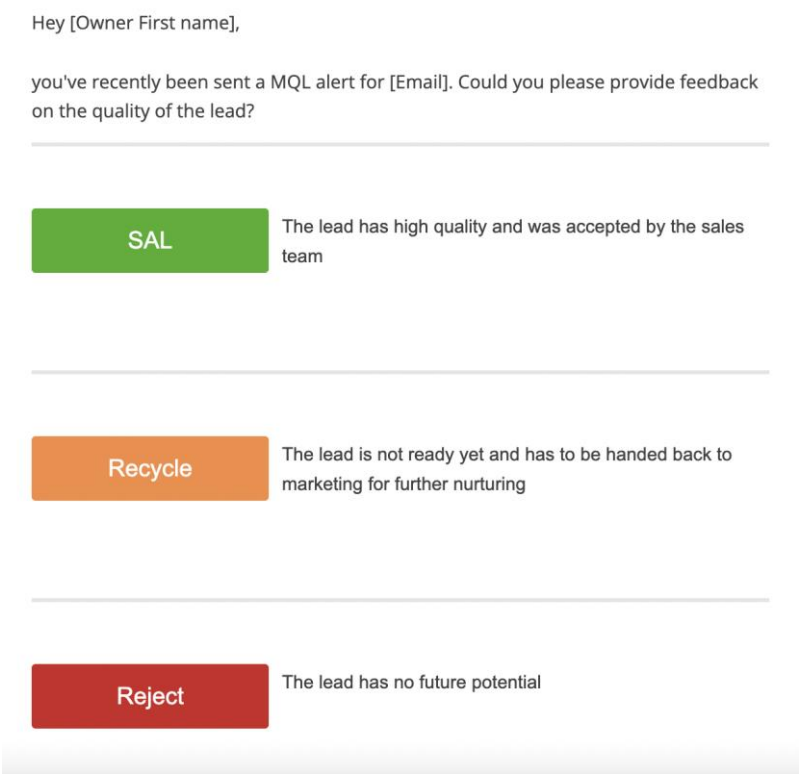

Abbildung 24: E-Mail zum Generieren von Vertriebsfeedback
Quelle: Eigene Darstellung

Dazu werden, je nachdem welcher der drei Buttons geklickt wird, unterschiedliche URL-Parameter an den Link einem Formular angehängt. URL-Parameter sind Zeichenketten, die an das Ende einer URL angehängt werden, um Informationen zu übermitteln. Die eigentliche URL wird dabei durch ein Fragezeichen von den Parametern getrennt (vgl. Sereda 2021). Sofern die Namen der einzelnen URL-Parameter mit den Feldnamen im Formular übereinstimmen, werden diese von Mautic automatisch in die Formularfelder übernommen. Beispielsweise wird in der folgenden URL die E-Mail-Adresse des Leads durch die Variable {contactfield=email} und das Feedback "SAL" in das Formular eingetragen.

„https://meinformular?email={contactfield=email}&salesfeedback=SAL“

Ist dies geschehen, so wird mithilfe des folgenden Java Scripts aus Abbildung 25 das Formular mit dem Vertriebsfeedback automatisch abgeschickt. Geht innerhalb von sieben Tagen kein Feedback vom Vertriebsteam ein, so wird die E-Mail erneut versandt. Dieser Vorgang wiederholt sich, bis Feedback zum MQL verfügbar ist.

```
<script>
  function autosubmit() {
    var form = document.querySelector('form[id="mauticform_autosubmitfeedbackform"]');
    if (form) {
      form.submit();
    }
  }
  window.addEventListener('load', autosubmit);
</script>
```

Abbildung 25: JavaScript-Code zum Abschicken des Vertriebsfeedbacks
Quelle: Eigene Darstellung

Anhang 2: Python-Code zur Aufbereitung der Lead-Daten

```
import os
import pandas as pd

# Import Mautic data
os.chdir('working_folder_path')
df1 = pd.read_csv('converted.csv')
df2 = pd.read_csv('not_converted.csv')

# Add column "Converted" and merge both DataFrames
df1['Converted'] = 1
df2['Converted'] = 0
df = pd.concat([df1, df2])

#Convert date columns into datetime
df['date_hit'] = pd.to_datetime(df['date_hit'])
df['end_date'] = pd.to_datetime(df['end_date'])
df['start_date'] = pd.to_datetime(df['start_date'])
df['confirmed_optout_date'] = pd.to_datetime(df['confirmed_optout_date'], errors='coerce')

# Delete pagehits and optouts which took time after the conversion
df = df[df['date_hit'] <= df['end_date']]
df.loc[df['confirmed_optout_date'] > df['end_date'], 'confirmed_optout_date'] = pd.NaT

# Replace empty optout dates with 0 und filled optout dates with 1
df['confirmed_optout_date'] = df['confirmed_optout_date'].apply(lambda x: 0 if pd.isnull(x) else 1)

#Calculate Sales Cycle Duration Column for unconverted and unconverted leads
df['time_diff'] = (df['end_date'] - df['start_date']).dt.total_seconds() / (24 * 60 * 60)

#Build Url Categories
df['url'] = df['url'].astype(str)
df.loc[df['url'].str.contains('Keyword1|Keyword2|Keyword3|...'), 'url'] = 'Product'
df.loc[df['url'].str.contains('Keyword1|Keyword2|Keyword3|...'), 'url'] = 'Checkout-Registration'
df.loc[df['url'].str.contains('Keyword1|Keyword2|Keyword3|...'), 'url'] = 'Registration'
df.loc[df['url'].str.contains('Keyword1|Keyword2|Keyword3|...'), 'url'] = 'Price'
df.loc[df['url'].str.contains('Keyword1|Keyword2|Keyword3|...'), 'url'] = 'Basket'
df.loc[df['url'].str.contains('Keyword1|Keyword2|Keyword3|...'), 'url'] = 'Checkout'
df.loc[df['url'].str.contains('Keyword1|Keyword2|Keyword3|...'), 'url'] = 'Contact'
df.loc[df['url'].str.contains('Keyword1|Keyword2|Keyword3|...'), 'url'] = 'Newsletter Registration'
df.loc[df['url'].str.contains('Keyword1|Keyword2|Keyword3|...'), 'url'] = 'Browsing 1'
df.loc[df['url'].str.contains('Keyword1|Keyword2|Keyword3|...'), 'url'] = 'Browsing 2'
df.loc[df['url'].str.contains('Keyword1|Keyword2|Keyword3|...'), 'url'] = 'Browsing 3'
df.loc[df['url'].str.contains('Keyword1|Keyword2|Keyword3|...'), 'url'] = 'Faq'
df.loc[df['url'].str.contains('Keyword1|Keyword2|Keyword3|...'), 'url'] = 'Account '
df = df[~df['url'].str.contains('confirm')]
df.loc[df['url'].str.contains('my_website'), 'url'] = 'Other Page'
```

```
df = df[~df['url'].str.contains('http')]



# Create DataFrame with consolidated data
#Count total pagehits for each contact
total_pagehits = df.groupby('lead_id')['url'].count().reset_index()
total_pagehits = total_pagehits.rename(columns={'url': 'total_pagehits'})
#Count number of pagehits in each category for each contact
pagehit_group = df.groupby(['lead_id', 'url']).size().unstack(fill_value=0).reset_index()
#Calculate salescycle duration for each contact
salescycle_duration = df.groupby('lead_id')['time_diff'].mean().round(2).reset_index()
salescycle_duration = salescycle_duration.rename(columns={'time_diff': 'salescycle_duration'})
#Collect conversion status for each contact
conversion_df = df.groupby('lead_id')['Converted'].mean().reset_index()
#Collect optout status for each lead
optout_df = df.groupby('lead_id')['confirmed_optout_date'].mean().reset_index()
optout_df = optout_df.rename(columns={'confirmed_optout_date': 'opted_out'})
#Combine the data in one DataFrame
merged_df = total_pagehits.merge(pagehit_group, on='lead_id', how='left')
merged_df = merged_df.merge(salescycle_duration, on='lead_id', how='left')
merged_df = merged_df.merge(optout_df, on='lead_id', how='left')
merged_df = merged_df.merge(conversion_df, on='lead_id', how='left')

# Remove entrys in which the number of total page visits is either less than 3 or higher than 300 (outliers)
merged_df = merged_df[(merged_df['total_pagehits'] <= 300) & (merged_df['total_pagehits'] >= 3)]
```

Abbildung 26: Python-Code zur Aufbereitung der Lead-Daten
Quelle: Eigene Darstellung

Anhang 3: SQL-Statements zum Download der Daten konvertierter Leads aus Mautic

```
SELECT
    lead_lists_leads.lead_id,
    lead_lists_leads.date_added AS end_date,
    page_hits.url,
    page_hits.date_hit,
    leads.date_added AS start_date,
    leads.confirmed_optout_date
FROM lead_lists_leads
JOIN leads ON lead_lists_leads.lead_id = leads.id
JOIN page_hits ON lead_lists_leads.lead_id = page_hits.lead_id
WHERE lead_lists_leads.leadlist_id = 4
  AND (leads.date_added LIKE '2021%' OR leads.date_added LIKE '2022%' OR leads.date_added LIKE '2023%')
  AND page_hits.url LIKE '%mywebsite%';
```

Abbildung 27: SQL-Statement zum Download der Daten konvertierter Leads aus Mautic
Quelle: Eigene Darstellung

Anhang 4: SQL-Statements zum Download der Daten nicht-konvertierter Leads aus Mautic

```
SELECT
    lead_lists_leads.lead_id,
    page_hits.url,
    page_hits.date_hit,
    leads.last_active AS end_date,
    leads.date_added AS start_date,
    leads.confirmed_optout_date
FROM leads
JOIN page_hits ON leads.id = page_hits.lead_id
JOIN lead_lists_leads ON leads.id = lead_lists_leads.lead_id
WHERE (leads.date_added LIKE '2021%' OR leads.date_added LIKE '2022%' OR leads.date_added LIKE '2023%')
  AND lead_lists_leads.leadlist_id IN (1, 2, 3)
  AND page_hits.url LIKE '%mywebsite%'
  AND NOT EXISTS (
      SELECT 1
      FROM lead_lists_leads
      WHERE lead_lists_leads.lead_id = leads.id
        AND lead_lists_leads.leadlist_id = 4
  );
```

Abbildung 28: SQL-Statement zum Download der Daten nicht-konvertierter Leads aus Mautic
Quelle: Eigene Darstellung

Anhang 5: Python-Code zur Analyse der Lead-Daten

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Split df into df_converted and df_unconverted
df_converted = merged_df[merged_df['Converted'] == 1]
df_unconverted = merged_df[merged_df['Converted'] == 0]

#Create df_stats
exclude_columns = ['lead_id', 'Converted', 'opted_out']
all_columns = df_converted.columns.difference(exclude_columns)
df_stats = pd.DataFrame(index=pd.MultiIndex.from_product([all_columns, ['Average', 'Stdev', 'Median', 'Max']]),
columns=['Converted', 'Unconverted'])
for col in all_columns:
    #Calculate stats
    avg_converted = df_converted[col].mean()
    stdev_converted = df_converted[col].std()
    median_converted = df_converted[col].median()
    max_converted = df_converted[col].max()
    avg_unconverted = df_unconverted[col].mean()
    stdev_unconverted = df_unconverted[col].std()
    median_unconverted = df_unconverted[col].median()
    max_unconverted = df_unconverted[col].max()
    # Round stats (2 digits)
    avg_converted = round(avg_converted, 2)
    stdev_converted = round(stdev_converted, 2)
    median_converted = round(median_converted, 2)
    max_converted = round(max_converted, 2)
    avg_unconverted = round(avg_unconverted, 2)
    stdev_unconverted = round(stdev_unconverted, 2)
    median_unconverted = round(median_unconverted, 2)
    max_unconverted = round(max_unconverted, 2)
    #Add to df_stats
    df_stats.at[(col, 'Average'), 'Converted'] = avg_converted
    df_stats.at[(col, 'Stdev'), 'Converted'] = stdev_converted
    df_stats.at[(col, 'Median'), 'Converted'] = median_converted
    df_stats.at[(col, 'Max'), 'Converted'] = max_converted
    df_stats.at[(col, 'Average'), 'Unconverted'] = avg_unconverted
    df_stats.at[(col, 'Stdev'), 'Unconverted'] = stdev_unconverted
    df_stats.at[(col, 'Median'), 'Unconverted'] = median_unconverted
    df_stats.at[(col, 'Max'), 'Unconverted'] = max_unconverted

#Add percentage opted out to df_stats
percentage_opted_out1 = (df_converted['opted_out'] == 1).mean() * 100
percentage_opted_out2 = (df_unconverted['opted_out'] == 1).mean() * 100
```

```
percentage_opted_out1 = round(percentage_opted_out1, 2)
percentage_opted_out2 = round(percentage_opted_out2, 2)
df_stats.at[('opted_out', 'Percentage opted out'), 'Converted'] = percentage_opted_out1
df_stats.at[('opted_out', 'Percentage opted out'), 'Unconverted'] = percentage_opted_out2

# Build correlation table
df_corr = merged_df.drop('lead_id', axis=1).copy()
correlation = df_corr.corr()
# Extract correlation with 'Converted'
correlation_with_converted = correlation['Converted'].to_frame()
# Create correlation heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_with_converted, annot=True, cmap='coolwarm', linewidths=.5)
plt.title('Correlation with "Converted"')
plt.show()
```

Abbildung 29: Python-Code zur Analyse der Lead-Daten
Quelle: Eigene Darstellung

Anhang 6: Python-Code zum Ermitteln und Anpassen des Schwellenwertes im Lead Scoring

```
import pandas as pd
import os
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import accuracy_score

# Import data
os.chdir('working_folder_path')
df = pd.read_excel('leadübersicht.xlsx')

# Define point system
column_weights = {
    'Basket': 10,
    'Browsing 1': 1,
    'Browsing 2': 1,
    'Browsing 3': 1,
    'Checkout Page': 25,
    'Checkout-Registration': 25,
    'Contact': 3,
    'Faq': 2,
    'MyAccount Page': 4,
    'Newsletter Registration': 3,
    'Other Page': 1,
    'Price Page': 5,
    'Product Page': 4,
    'Registration': 3,
    'opted_out': -25
}

# Calculate score for each lead
df['weighted_sum'] = df[list(column_weights.keys())].multiply(list(column_weights.values())).sum(axis=1)

# Create df for converted and nonconverted lead scores
converted_1 = df[df['Converted'] == 1]['weighted_sum']
converted_0 = df[df['Converted'] == 0]['weighted_sum']

# Create histogram for converted and non converted lead scores
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(converted_1, bins=range(-50, 600, 10), edgecolor='black')
plt.title('Histogram for converted leads')
plt.xlabel('Score')
plt.ylabel('Number of leads')
plt.subplot(1, 2, 2)
```

```
plt.hist(converted_0, bins=range(-50, 600, 10), edgecolor='black')
plt.title('Histogram for unconverted Leads')
plt.xlabel('Score')
plt.ylabel('Number of leads')
plt.tight_layout()
plt.show()



# Create percentile table
percentiles = np.arange(0, 1, 0.05)
# For nonconverted leads
table_converted_0 = pd.DataFrame({
    'Percentile': percentiles,
    'Value': [df[df['Converted'] == 0]['weighted_sum'].quantile(p) for p in percentiles]
})
# For converted leads
table_converted_1 = pd.DataFrame({
    'Percentile': percentiles,
    'Value': [df[df['Converted'] == 1]['weighted_sum'].quantile(p) for p in percentiles]
})
# Combined table
table_combined = pd.DataFrame({
    'Percentile': percentiles,
    'Converted_0': [df[df['Converted'] == 0]['weighted_sum'].quantile(p) for p in percentiles],
    'Converted_1': [df[df['Converted'] == 1]['weighted_sum'].quantile(p) for p in percentiles]
})
# Print combined percentile table
print("\nKombinierte Tabelle:")
print(table_combined)

# Define treshold
threshold = 52

# Create 'Prediction' column
df['Prediction'] = np.where(df['weighted_sum'] > threshold, 1, 0)

# Create heatmap
confusion_matrix = pd.crosstab(df['Converted'], df['Prediction'], rownames=['Actual'], colnames=['Predicted'])
sns.heatmap(confusion_matrix, annot=True, fmt='d', cmap='Blues', cbar=False)
plt.show()

# Calculate accuracy for choosen treshold
accuracy = accuracy_score(df['Converted'], df['Prediction'])
print(f'Accuracy: {accuracy}')
```

Abbildung 30: Python-Code zum Ermitteln und Anpassen des Schwellenwertes im Lead Scoring
Quelle: Eigene Darstellung

Anhang 7: Ergebnisse des prädiktiven Lead Scorings mit dem Support-Vector-Machine-Algorithmus

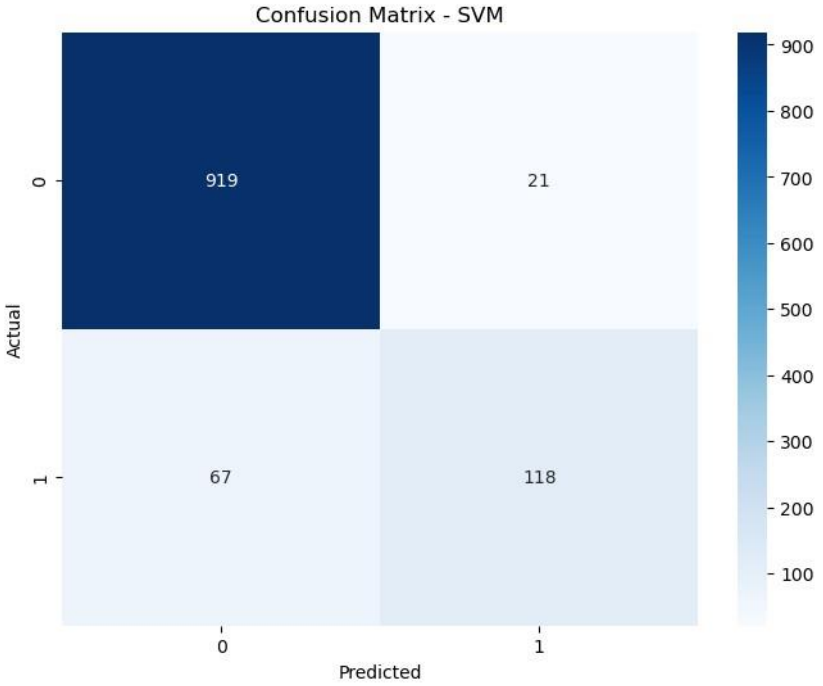

Abbildung 31: Ergebnisse des prädiktiven Lead Scorings mit dem Support-Vector-Machine-Algorithmus
Quelle: Eigene Darstellung

Anhang 8: Ergebnisse des prädiktiven Lead Scorings mit dem Decision-Tree-Algorithmus

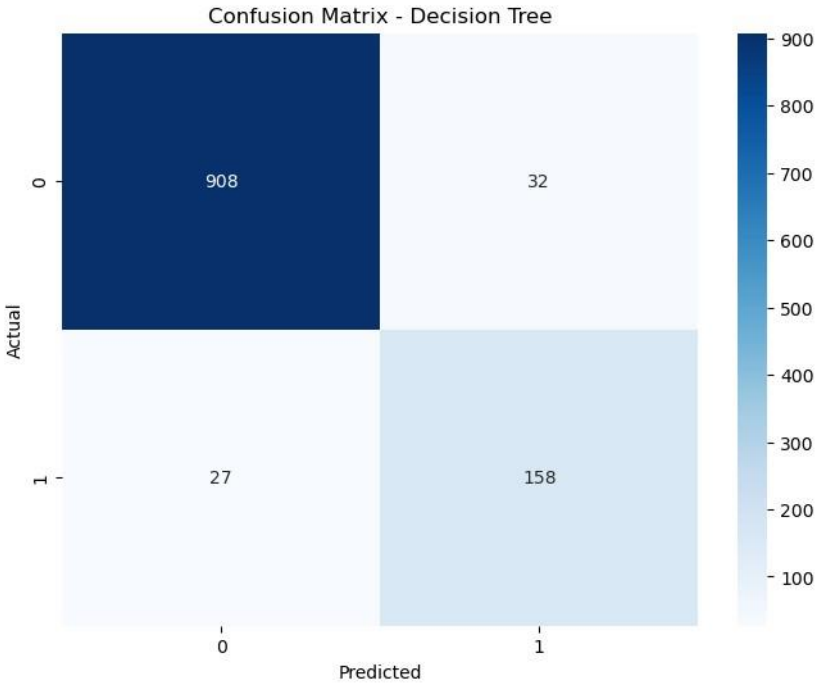

Abbildung 32: Ergebnisse des prädiktiven Lead Scorings mit dem Decision-Tree-Algorithmus
Quelle: Eigene Darstellung

Anhang 9: Ergebnisse des prädiktiven Lead Scorings mit dem Logistic-Regression-Algorithmus

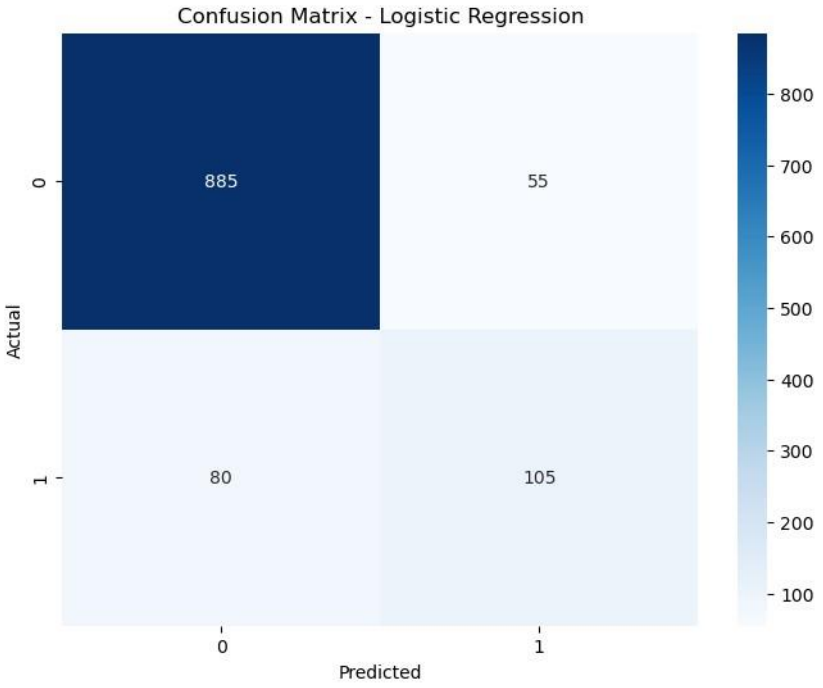

Abbildung 33: Ergebnisse des prädiktiven Lead Scorings mit dem Logistic-Regression-Algorithmus
Quelle: Eigene Darstellung

Anhang 10: Python-Code zum Entwickeln eines Machine-Learning-Modells für das Lead Scoring

```
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, confusion_matrix
import joblib

# Import data
os.chdir('working_folder_path')
df = pd.read_excel('leadübersicht.xlsx')

# Split into features (X) and target variable (y)
X = df.drop(["Converted", "lead_id"], axis=1)
y = df["Converted"]

# Scaling and splitting into training and test data
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)

# Define models and hyperparameters
models = {
  'Logistic Regression': (LogisticRegression(), {'C': [0.1, 1, 10]}),
   'Decision Tree': (DecisionTreeClassifier(), {'max_depth': [None, 10, 20, 30]}),
   'Random Forest': (RandomForestClassifier(), {'n_estimators': [50, 100, 200]}),
   'SVM': (SVC(), {'C': [0.1, 1, 10], 'kernel': ['linear', 'rbf']})
}

# Executing GridSearchCV for every model
best_model = None
best_accuracy = 0.0

for name, (model, params) in models.items():
   grid_search = GridSearchCV(model, params, cv=5, scoring='accuracy')
   grid_search.fit(X_train, y_train)

   #Displaying the best hyperparameters
   print(f"Best parameters for {name}: {grid_search.best_params_}")
```

```
    # Evaluation on the test data
    y_pred = grid_search.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    print(f"Accuracy for {name}: {accuracy}")

    if accuracy > best_accuracy:
        best_model = grid_search.best_estimator_
        best_accuracy = accuracy

    #Confusion Matrix
    cm = confusion_matrix(y_test, y_pred)
    plt.figure(figsize=(8, 6))
    sns.heatmap(cm, annot=True, fmt="d", cmap="Blues", xticklabels=['0', '1'], yticklabels=['0', '1'])
    plt.title(f"Confusion Matrix - {name}")
    plt.xlabel("Predicted")
    plt.ylabel("Actual")
    # Saving plots as png
    plot_filename = f"{name}_confusion_matrix.png"
    plt.savefig(plot_filename, format='png', bbox_inches='tight')

#Save best model and used scaler
joblib.dump(scaler, 'scaler.joblib')
joblib.dump(best_model, 'best_model.joblib')
```

Abbildung 34: Python-Code zum Entwickeln eines Machine-Learning-Modells für das Lead Scoring
Quelle: Eigene Darstellung

Anhang 11: Python-Code zur Anwendung des Machine-Learning-Modells

```python
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
import joblib
import os

# Load model and scaler
os.chdir('working_folder_path')
loaded_model = joblib.load('best_model.joblib')
scaler = joblib.load('scaler.joblib')

# Load the data on which a prediction has to be made
df_new = pd.read_excel('leadübersicht.xlsx')
X_new = df_new.drop(["Converted", "lead_id"], axis=1)
X_new_scaled = scaler.transform(X_new)

# Make predictions
predictions = loaded_model.predict(X_new_scaled)

# Create df to be exported
df_predictions = pd.DataFrame({
    'lead_id': df_new['lead_id'],
    'Predicted_Converted': predictions
})

# Export df as csv
df_predictions.to_csv('new_predictions.csv', index=False)
```

Abbildung 35: Python-Code zur Anwendung des Machine-Learning-Modells
Quelle: Eigene Darstellung