MODELLE IM VERGLEICH

Wie wichtig ist die Auswahl des Algorithmus?

Ihre Herausforderung:

- ? Welche alternativen Algorithmen zum Scoring gibt es?
- ? Wie ist die Trennschärfe beim Scoring für die einzelnen Algorithmen einzuschätzen?
- ? Was sind die Vor- und Nachteile von Regressionsmethoden beim Scoring?

Unsere Lösung:

Klassische Methoden wie lineare oder logistische Regression sind bewährte Verfahren, um Scorings im Marketing zu erstellen. Liefern alternative Verfahren wie Random Forest oder neuronale Netze im Kontext von Marketing-Scorings tatsächlich bessere Ergebnisse als die klassischen Regressionsmethoden? Wenn ja, wie groß ist dieser Unterschied? Wie sieht es aus mit Rechner-Laufzeit und der Interpretierbarkeit bei der Verwendung eines Black-Box-Verfahrens aus? Wir geben Ihnen Antworten auf diese Fragen.

Ihr Nutzen:

- ✓ Sie erhalten einen **Überblick** über die möglichen Alternativen.
- ✓ Sie können die verschiedenen **Algorithmen besser bewerten**.
- ✓ Sie können einen für Ihre Zwecke **geeigneten Algorithmus auswählen**.

Sie wollen mehr wissen? Sprechen Sie uns an!

Modelle im Vergleich – wie wichtig ist die Auswahl des Algorithmus?

mar,an,con unterstützt Unternehmen mit Scorings für verschiedene Anwendungsfälle. Interessenten-, Cross-Selling-, Wiederkauf-, Churn-Prevention oder Reaktivierungs-Score dienen dazu, in der betrefenden Situation die jeweils attraktivsten Kunden zu identifizieren. Im Reaktivierungs-Score wird beispielsweise eine Reaktivierungswahrscheinlichkeit für jeden Kunden ermittelt. So lassen sich die Kunden mit den größten Reaktivierungsaussichten gezielt ansprechen. Je trennschärfer das verwendete Scoring aussichtsreiche von weniger aussichtsreichen Kunden unterscheiden kann, umso besser. Neben der richtigen Variablenauswahl und -aufbereitung kann dabei auch die richtige Auswahl des Scoring-Algorithmus relevant sein. Bisher verwendet mar,an,con meist die logistische Regression, ein bewährtes, stabil laufendes Verfahren mit vergleichsweise kurzen Rechenzeiten. Im Folgenden werden an einem typischen Datenbeispiel einige Alternativen der Künstlichen Intelligenz getestet, um zu sehen, inwieweit sie bessere Ergebnisse liefern können.

Viele Algorithmen – ein kurzer Überblick

Im Folgenden sollen die verwendeten Algorithmen kurz beschrieben werden, ohne näher auf technische Details einzugehen. Alle Algorithmen stehen vor der Aufgabe, „ja“ und „nein“, beispielsweise Reaktivierung oder Nicht-Reaktivierung möglichst gut vorherzusagen. Die Qualität dieser Vorhersage lässt sich mit der weiter unten beschriebenen Area under Curve (AUC) messen.

Logistische Regession: Hier wird eine gewichtete Summe der Einflussgrößen bestimmt, der sog. lineare Prediktor. Die geschätzte Wahrscheinlichkeit ergibt sich dann über die logistische Funktion, eine S-förmige Funktion, aus dem linearen Prediktor. Der Algorithmus muss diejenige Gewichtung der Einflussgrößen bestimmen, mit denen sich eine möglichst gute Vorhersage der Reaktivierungswahrscheinlichkeit ergibt.

Entscheidungsbaum: Hierbei werden die Kunden anhand eines Merkmals in zwei Gruppen unterteilt: eine mit höherem und eine mit niedrigerem Reaktivierungsanteil. Wendet man für die so definierten Gruppen dieses Vorgehen erneut an, entstehen Untergruppen und bei wiederholter Anwendung ein sich immer weiter verzweigender Baum. Am Ende der Äste des Baums stehen dann jeweils mehr oder weniger homogene Gruppen mit einer bestimmten Reaktivierungswahrscheinlichkeit.

Random Forest: Aufbauend auf dem Entscheidungsbaum-Verfahren werden hier zufällige „Baumstümpfe“ erzeugt: Auf einer jeweils zufällig ausgewählten Teilmenge der Kunden wird ein Entscheidungsbaum erzeugt, wobei an jeder Verzweigung nur ein zufällig ausgewählter Teil der Einflussmerkmale verwendet wird. Auf diese Weise entsteht eine Vielzahl ähnlicher Bäume, und die Reaktivierungswahrscheinlichkeit eines Kunden ergibt sich aus seiner mittleren Wahrscheinlichkeit in den einzelnen Bäumen.

Extra Trees: Auch hier werden zufällige Entscheidungsbäume erzeugt. Dabei werden jeweils alle zur Verfügung stehenden Kunden verwendet, jedoch steht an den Verzweigungen immer nur ein zufälliger Teil der Merkmale zur Verfügung, wobei der bei den Merkmalen verwendete Trennwert zufällig festgelegt wird. Ebenso wie beim Random Forest werden hier zahlreiche suboptimale Bäume erzeugt, die aber in ihrer Masse dann zu sinnvollen Ergebnissen führen.

Nächste Nachbarn: Hier wird für jeden Kunden geschaut, wie sich seine nächsten Nachbarn verhalten, also die Kunden mit den ähnlichsten Werten bei den Einflussgrößen. Der Anteil der nächsten Nachbarn, die reaktiviert wurden, ergibt dann die Reaktivierungswahrscheinlichkeit.

Multi-Layer Perceptron Classifier (MLP): Dieses Verfahren baut neuronale Netze auf. Dabei werden ähnlich wie bei der logistischen Regression gewichtete Summen der Merkmale gebildet und anschließend über eine S-förmige Funktion zu einem latenten Merkmal („Neuron“) transformiert. In einer zweiten Ebene geschieht mit den Neuronen dasselbe, bis schließlich die Neuronen der letzten Ebene auf gleiche Weise zur Reaktionswahrscheinlichkeit kombiniert werden. Auf diese Weise lassen sich je nach Zahl der Ebenen und Neuronen der Ebene hochkomplexe nichtlineare Zusammenhänge modellieren.

Adaptive Boosting: Es werden einfache Zuordnungsregeln zu Reaktivierten bzw. nicht Reaktivierten gesucht, die nur geringfügig besser als eine zufällige Einteilung sein müssen. Diese betrachten z.B. oft nur ein einzelnes Merkmal oder Kombi-

nationen zweier Merkmale. Das adaptive Boosting entwickelt nun aus diesen einfachen Klassifikatoren eine übergeordnete Zuordnungsregel, indem es diese Zuordnungsergebnisse gewichtet aggregiert. Die finale Zuordnung ergibt sich dann aus der gewichteten Mehrheit der Einzelzuordnungen für eine zu bestimmende optimale Gewichtung.

Naive Bayes-Klassifikation: Für jeden möglichen Merkmalswert jeder Einflussgröße wird das Verhältnis aus Reagierern und Nicht-Reagieren bestimmt, sodass sich jeweils ein für diesen Merkmalswert spezifisches Chancenverhältnis der Reaktivierung ergibt. Das Chancenverhältnis des Kunden ergibt sich dann als geometrisches Mittel der Chancenverhältnisse seiner Merkmalswerte.

Diskriminanzanalyse: Die Zuordnung zu Reaktivierten oder Nicht-Reaktivierten ergibt sich durch eine trennende Ebene. Die Reaktivierungswahrscheinlichkeit ergibt sich dann aus der Entfernung zu dieser trennenden Ebene. Bei der quadratischen Diskriminanzanalyse (QDA) wird statt der Ebene eine gebogene Fläche verwednet, die sich aus einer quadratischen Funktion der Einflussmerkmale ergibt.

Vergleichskriterien – worauf kommt es bei der Anwendung an?

Möglichst gute Prognosequalität allein ist nicht alles. Es kommt auch auf den Aufwand an, mit dem diese Prognosequalität erreicht wird. Und auch auf die Frage, wie gut sich die Ergebnisse auch für den analytischen Laien beurteilen und interpretieren lassen. Im Folgenden werden daher die folgenden Kriterien beim Vergleich der Algorithmen herangezogen:

Prognosequalität des Scores: Wie gut sind die berechneten Wahrscheinlichkeiten geeignet, die reaktivierbaren Kunden vorauszusagen? Um dies zu beurteilen, kann die sog. Area under Curve verwendet werden (vgl. Abb. 1). Dazu sortiert man die Kunden nach im Scorewert absteigend. Dann sollten zunächst verhältnismäßig viele Reagierer auftreten, im weiteren Verlauf immer weniger.

Trägt man nun in dieser Sortierung den kumulierten Kundenanteil und den kumulierten Anteil der Reagierer ab, so ergibt sich im Idealfall eine schnell ansteigende Kurve, die später immer weiter abflacht. Der Score funktioniert gut, wenn z.B. auf 25% der Kunden mit den besten Scorewerten bereits 75% der Reagierer entfallen. Um die Qualität insgesamt zu beurteilen, kann in dem Quadrat der kumulierten Kunden und Reagierer die Fläche unter der Kurve betrachtet werden. Ein AUC-Wert von 0,7 ist hierbei schon sehr gut. Funktioniert der Score überhaupt nicht, treten die Reagierer in der Score-Sortierung rein zufällig auf. Die Kurve entspricht dann der Diagonalen und es gilt AUC=0,5.

Rechenzeit: Die Algorithmen unterscheiden sich deutlich in der benötigten Rechenzeit. Je mehr

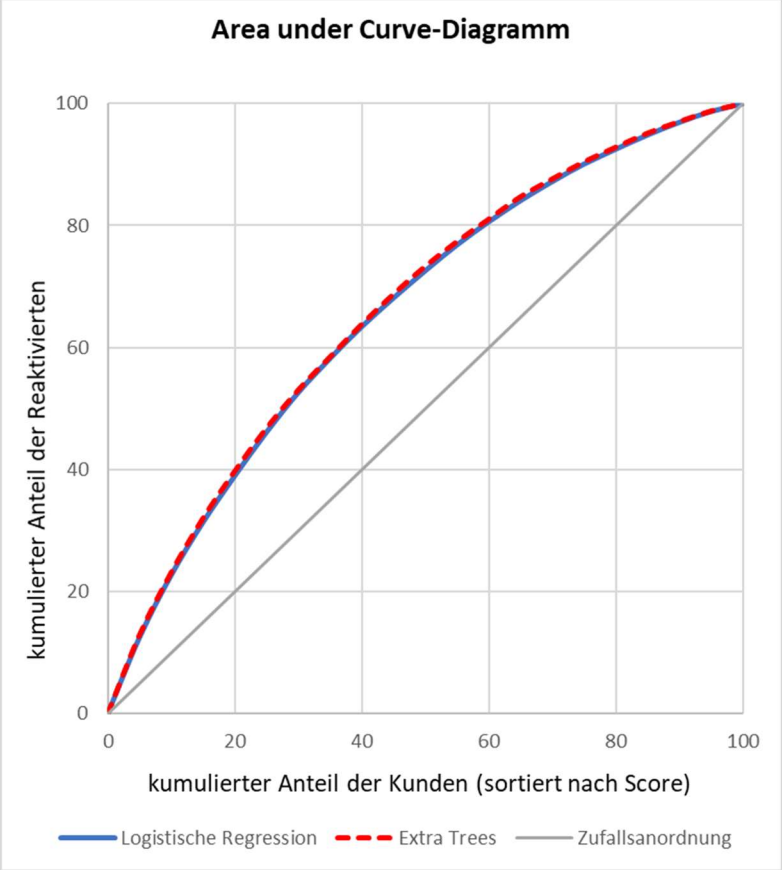

Abbildung 1: *Grafische Darstellung der Area under Curve. Der AUC-Wert stellt die Fläche zwischen x-Achse und der Kurve dar. Benchmark ist die zufällige Anordnung der Kunden; in diesem Fall ist die Kurve eine Diagonale und AUC = 0,5. Hier dargestellt sind das Scoring mit logistischer Regression (AUC = 0,661) und mit Extra Trees (AUC = 0,666), jeweils für transformierte Eingangsdaten.*

Freiheiten das Verfahren bekommt, umso mehr Rechenzeit wird tendenziell benötigt. Meist müssen auch noch verschiedene Rahmenbedingungen ausgetestet werden, z.B. bei MLP-Classifiern die Anzahl und Größe der zu verwendenden latenten Neuronen-Ebenen. Bei einem einzigen Rechenlauf wäre die Rechenzeit prinzipiell nicht so wichtig. Da man aber häufig verschiedene Varianten der Datenauswahl und -aufbereitung testen muss, macht sich eine lange Rechenzeit dann doch störend bemerkbar.

Interpretierbarkeit: Für die Akzeptanz des Scores unter Anwendern ist sehr hilfreich, wenn sein Ergebnis einleuchtet und zu bisherigen Erfahrungswerten passt. In dieser Beziehung ist die logistische Regression ideal: Sie liefert Parameterwerte β, deren Exponentialwerte e^{β} als relative Veränderung der Reaktivierungschance bei Veränderung des Einflussmerkmals interpretierbar sind. Demgegenüber sind die meisten anderen Verfahren eine „Black Box“: Ohne weitere Zusatzanalysen wird nicht ersichtlich, wie die Einflussgrößen sich auf die berechnete Wahrscheinlichkeit auswirken. So bleiben z.B. bei MLP-Classifiern die dazwischen liegenden Ebenen der Neuronen im Verborgenen, sodass nicht klar ist, wie die Einflussgrößen und die Zielgröße zusammenhängen.

Die Algorithmen im Test – welche Rechenmethode trennt am besten?

Um die insgesamt neun verschiedenen Algorithmen zu vergleichen, wurden diese auf Testdaten angewendet. Dabei handelt es sich um Daten eines Filialisten, wobei das Ziel war, unter Kunden mit einer Inaktivität seit mehr als zwei Jahren diejenigen mit der größten Reaktivierungswahrscheinlichkeit zu finden. Während in der realen Analyse mehr als 50 Einflussmerkmale betrachtet wurden, betrachtet der Beispieldatensatz davon nur sechs besonders wichtige, die auch unabhängig von dem konkreten Unternehmen einleuchten:

- Recency: Wie lange ist es her, dass der Kunde zuletzt gekauft hat?
- Frequency: Wie oft hat der Kunde schon gekauft?
- Monetary Value: Wie hoch ist der bisherige Gesamtumsatz des Kunden?
- Alter: Wie alt ist der Kunde?
- Haushaltscluster: Handelt es sich um einen Einzelkunden oder sind mehrere Kunden aus dem gleichen Haushalt in der Kundendatenbank vorhanden?

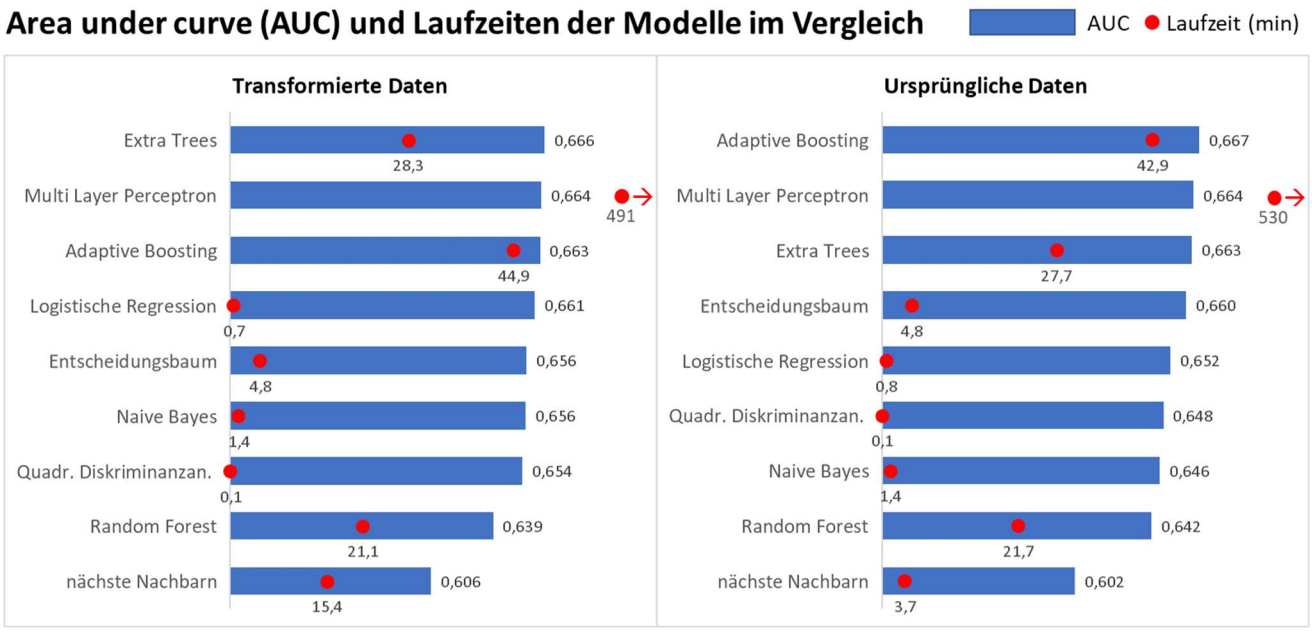

Abbildung 2: *Die Logistische Regression weist sehr kurze Laufzeiten auf und kann insbesondere bei der Verwendung der transformierten Daten (links) gut mit den anderen Verfahren mithalten. Die dargestellten Laufzeiten beziehen sich jeweils auf die vollständige Gittersuche, die AUC ist diejenige des jeweils besten Gitterdurchlaufs.*

- Kontakte: Wie häufig wurde der Kunde in den letzten zwei Jahren per Brief kontaktiert?

Für die traditionelle logistische Regression hat es sich als hilfreich erwiesen, diese Merkmale zum Teil zu transformieren. Die RFM-Merkmale (Recency, Frequncy, Monetary Value) sowie die Kontakanzahl wurden zu diesem Zweck logarithmiert, beim Alter wurde die Altersdifferenz zu 60 Jahren betrachtet. Hierbei nutzt der Analyst Erfahrungswerte und Eindrücke aus der Sicht auf die Daten, um die Eingangsdaten möglichst gut an die Linearitätsanforderungen der logistischen Regression anzupassen und insbesondere z.B. den Einfluss extrem großer Werte bei den Einflussgrößen zu normalisieren. Bei anderen, nicht linearen Verfahren ist zu erwarten, dass dies weniger relevant ist, diese also auch mit den nicht transformierten Originalwerten der sechs Einflussmerkmale zurechtkommen. Im Rahmen dieses Tests wurden daher die Analyseergebnisse je einmal mit den Originaldaten und mit den transformierten Daten verglichen.

Aus der Zahl aller inaktiven Kunden zu einem bestimmten Zeitpunkt fließt eine Stichprobe von 1.000.000 Kunden in die Analyse ein. Als Zielgröße wurde die Reaktivierung dieser Kunden innerhalb eines Jahres betrachtet, die möglichst gut vorherzusagen war.

Diese Stichprobe wurde dann nochmal in eine Lern- und eine Kontrollstichprobe unterteilt. Die von dem jeweiligen Scoring-Algorithmus mit 700.000 Kunden errechneten Reaktivierungswahrscheinlichkeiten bzw. Scores sollten sich dann an den weiteren 300.000 Kunden bewähren. Damit soll eine Überanpassung (Overfitting) ausgeschlossen werden, die dadurch entstehen kann, dass ein Verfahren mit zu vielen Freiheitsgraden die Besonderheiten der Lernstichprobe hervorragend erklären kann, sich das dann aber nicht auf die Kontrollgruppe übertragen lässt.

Bei den meisten Verfahren sind zudem weitere Rahmenparameter zu setzen, die sich auf die Qualität des Scoring und insbesondere ein eventuelles Overfitting auswirken, beim MLP-Classifier z.B. die Zahl der Neuronen-Ebenen und die Zahl der Neuronen pro Ebene. Um hier eine optimale Konstellation zu finden, durchläuft der Test eine Gittersuche. Dabei wird die Konstellation mit dem besten Scoring-Ergebnis ausgewählt, gemessen an der Area under Curve (AUC) für die Kontrollgruppe.

Abb. 2 zeigt in der linken Hälfte das Ergebnis dieses Tests für die beschriebenen Algorithmen anhand der transformierten Daten. Es zeigt sich, dass drei

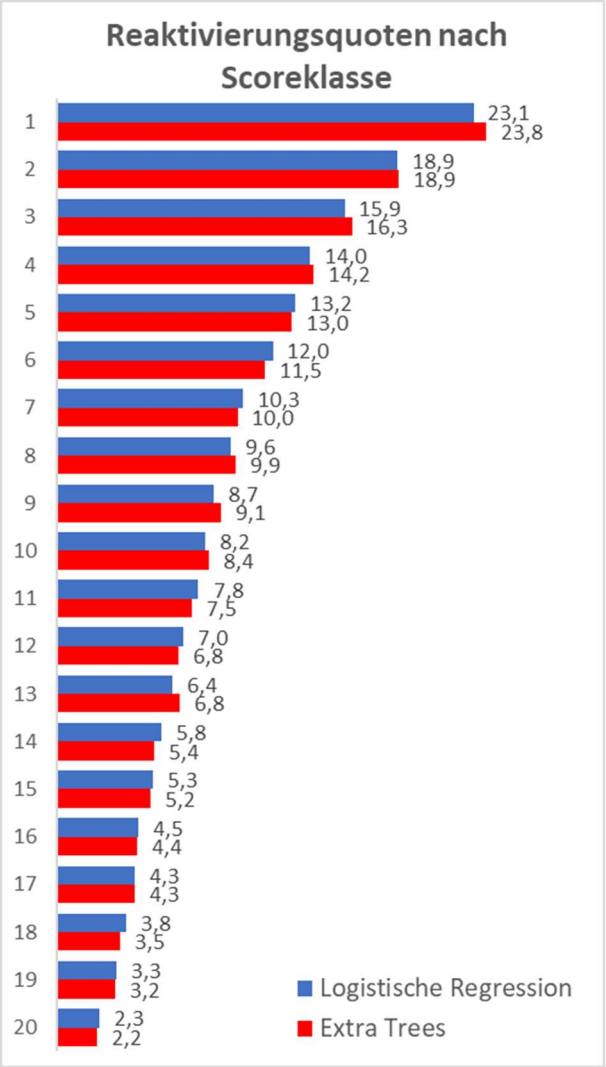

Abbildung 3: *Reaktivierungsquoten für zwei Algorithmen im Vergleich für die transformierten Daten.*

Algorithmen bessere AUC-Werte als die logistische Regression liefern: Extra Trees (0,666), MLP (0,664) und Adaptive Boosting (0,663). Dabei hat Adaptive Boosting allerdings den Nachteil, dass sich die Score-Werte nicht als Reaktivierungswahrscheinlichkeiten interpretieren lassen. Als nächstes Verfahren folgt die logistische Regression mit 0,661. Weitere drei Verfahren erreichen noch AUC-Werte über 0,65, ganz am Ende rangiert Nächste Nachbarn mit einem AUC-Wert von nur 0,606.

Der Genauigkeitsgewinn wird mit einer erheblich längeren Laufzeit erkauft. Bei den Laufzeiten für die gesamte Gittersuche liegt die logistische Regression mit 0,7 Minuten auf Rang 2, was nur noch durch die quadratische Dsikriminanzanalyse mit 0,1 Minuten unterboten wird – hier ist aber der AUC-Wert schon deutlich schlechter. Die beim AUC-Kriterium besten Verfahren benötigen 28 min (Extra Trees), 491 min (MLP) bzw. 45 min (Adaptive Boosting).

Kreuztabelle für Scorings mit Logistischer Regression und Extra Trees (gemeinsame Häufigkeiten in %)																					
		Scoreklasse für Logistische Regression																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Summe
Scoreklasse für Extra Trees	1	4,31	0,65	0,03	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5,00
	2	0,65	3,21	0,97	0,13	0,02	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5,00
	3	0,02	0,92	2,55	1,17	0,26	0,06	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5,00
	4	0,00	0,19	1,00	2,16	1,19	0,33	0,09	0,02	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5,00
	5	0,00	0,02	0,33	0,90	1,94	1,25	0,40	0,12	0,04	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5,00
	6	0,00	0,00	0,08	0,41	0,86	1,68	1,30	0,42	0,16	0,05	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5,00
	7	0,00	0,00	0,03	0,11	0,40	0,88	1,59	1,27	0,48	0,16	0,05	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5,00
	8	0,00	0,00	0,01	0,08	0,14	0,33	0,79	1,51	1,30	0,58	0,17	0,06	0,02	0,01	0,00	0,00	0,00	0,00	0,00	0,00	5,00
	9	0,00	0,00	0,00	0,03	0,14	0,23	0,29	0,71	1,41	1,29	0,65	0,17	0,05	0,01	0,00	0,00	0,00	0,00	0,00	0,00	5,00
	10	0,00	0,00	0,00	0,00	0,04	0,17	0,32	0,41	0,62	1,28	1,24	0,67	0,18	0,04	0,01	0,00	0,00	0,00	0,00	0,00	5,00
	11	0,00	0,00	0,00	0,00	0,01	0,04	0,14	0,34	0,54	0,66	1,16	1,15	0,70	0,22	0,03	0,01	0,00	0,00	0,00	0,00	5,00
	12	0,00	0,00	0,00	0,00	0,00	0,01	0,04	0,12	0,28	0,57	0,82	1,12	1,12	0,68	0,20	0,04	0,01	0,00	0,00	0,00	5,00
	13	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,04	0,10	0,22	0,55	0,99	1,12	1,02	0,69	0,20	0,04	0,00	0,00	0,00	5,00
	14	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,04	0,09	0,19	0,54	1,09	1,22	0,97	0,63	0,18	0,02	0,00	0,00	5,00
	15	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,02	0,04	0,07	0,14	0,48	1,19	1,43	0,94	0,51	0,16	0,01	0,00	5,00
	16	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,02	0,04	0,06	0,11	0,35	1,16	1,66	1,08	0,43	0,07	0,00	5,00
	17	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,02	0,04	0,07	0,13	0,29	1,06	1,82	1,20	0,32	0,02	5,00
	18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,02	0,04	0,08	0,12	0,30	1,05	1,92	1,31	0,13	5,00
	19	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,02	0,04	0,08	0,14	0,30	1,20	2,14	1,07	5,00
	20	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,06	1,14	3,77	5,00
	Summe	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	100,00

***Abbildung 4**: Kreuztabellierung von Scorings mit logistischer Regression und Extra Trees für transformierte Daten. Obwohl die AUCs beider Methoden sehr ähnlich sind, ergeben sich doch z.T. deutliche Abweichungen in der Scoreklassen-Zuordnung. Abweichungen von mehr als einer Scoreklassen sind durchaus häufig (insgesamt 22% aller Kunden).*

Dasselbe Vorgehen wurde auch noch mal für die originalen, nicht transformierten Daten durchgeführt. Die Erwartung war, dass die logistische Regression, die ein lineares Wirken der Einflussmerkmale auf die Zielgröße unterstellt, stärker abfällt, wenn die Daten nicht in optimaler Weise für das Verfahren aufbereitet wurden. Dies hat sich auch bewahrheitet, wie in Abb. 2 rechts dargestellt: Am besten schneidet hier Adaptive Boosting ab (AUC=0,667) vor MLP (0,664) und Extra Trees (0,663). Die logistische Regression folgt mit AUC=0,652 erst auf Rang 5; auf Rang 4 liegt nun der Entscheidungsbaum. Insgesamt schneidet die logistische Regression jedoch nicht so viel schlechter ab, als a priori befürchtet wurde.

Die Laufzeiten verhalten sich ähnlich wie bei den transformierten Daten: Auch hier liefert die logistische Regression nach 0,8 Minuten deutlich schneller Ergebnisse als die drei besten Verfahren; der Entscheidungsbaum hat auch noch eine verhältnismäßig kurze Laufzeit.

Der Unterschied zwischen den AUC-Werten zweier Algorithmen sieht auf den ersten Blick geringfügig aus. Um dies genauer einzuschätzen, kann man die Scoring-Ergebnisse detaillierter untersuchen. Dies geschieht im Folgenden für das Extra Trees-Scoring im Vergleich zur logistischen Regression. Teilt man die Kunden nach den Scores in jeweils 20 gleich große Klassen ein, so kann man zunächst die Anteile an Reaktivierten in diesen Klassen

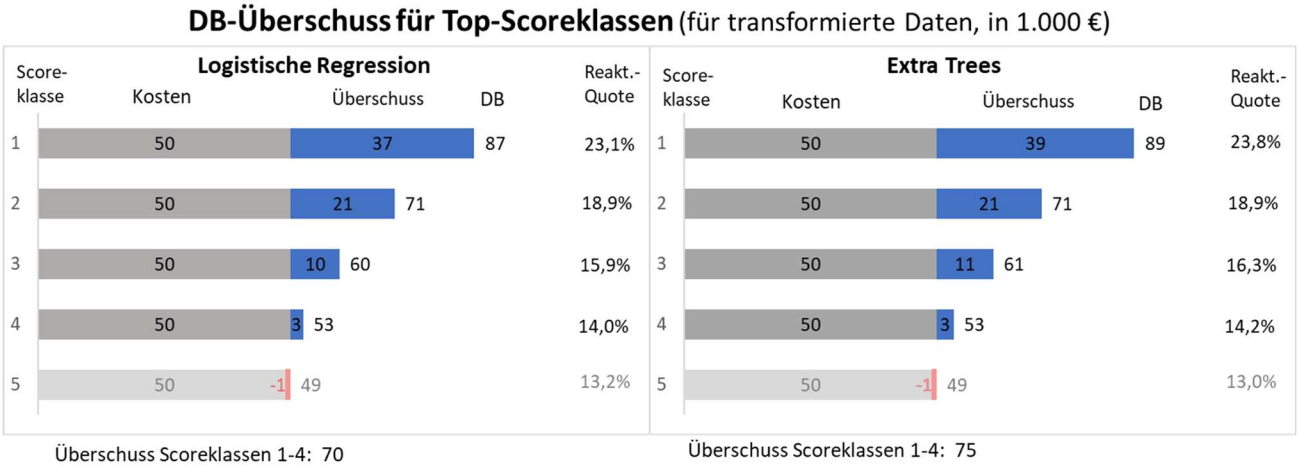

***Abbildung 5:** Überschuss des Deckungsbeitrags (DB) über die Mailingkosten bei Bewerbung der jeweils ersten vier Scoreklassen (Annahme: Mailingkosten 1€, Deckungsbeitrag pro Reaktivierten: 7,5 €) für die logistische Regression und Extra Trees. Der Überschuss liegt für Extra Trees mit ca. 75.000 € um ca. 5.000 € (+7,5%) über demjenigen der logistischen Regression.*

untersuchen. Diese fallen monoton mit der Scoreklasse, von 23,8% auf 2,2% bei Extra Trees und von 23,1% auf 2,3% für die logistische Regression (vgl. Abb. 3). Auch hier sind die Unterscheide sehr gering.
Trägt man aber die Scoreklassen-Einteilungen beider Algorithmen in einer Kreuztabelle gegeneinander ab, zeigen sich doch deutliche Unterschiede (vgl. Abb. 4). Die Algorithmen sind sich im Einzelfall nicht darüber einig, welche Kunden zu den am besten reaktivierbaren Kunden zu zählen sind, Abweichungen von mehr als drei Scoreklassen sind keine Seltenheit. Dies kann sich auch durchaus kostenmäßig auswirken. Als Modellrechnung ist hier unterstellt, dass das Kontaktieren eines Kunden 1,00 € kostet und dass bei einer Reaktivierung mit einem mittleren Deckungsbeitrag von 7,50 € zu rechnen ist – unabhängig von der Scoreklasse des Reaktivierten. Mit diesen Annahmen lässt sich z.B. berechnen, welcher Deckungsbeitrag sich nach Abzug der Kommunikationskosten erzielen lässt, wenn man die jeweils besten vier Scoreklassen kontaktiert. Das Ergebnis zeigt Abb. 5: In beiden Fällen liefern die besten vier Scoreklassen positive Überschüsse der Deckungsbeiträge über die Werbungskosten. Mit der logistischen Regression kontaktiert man damit 35.943 Reaktivierte und erwirtschaftet damit 69.600 €; mit Extra Trees erreicht man 36.653 Reaktivierte und erwirtschaftet 74.900 €. Dieser Unterschied ist dann doch schon größer als nur marginal.

Fazit – Was kann man für die Analysepraxis lernen?

Insgesamt zeigt sich, dass die logistische Regression zwar nicht optimal ist, aber andererseits der Unterscheid zu den besseren Verfahren nicht allzu groß ist. Dies gilt umso mehr, je mehr Zeit der Datenanalyst vorher in eine passende Datenaufbereitung investiert.
Datentransformationen ermöglichen es, eine Situation herzustellen, die den Modellannnahmen der logistischen Regression entspricht. So ist es plausibel, dass die Veränderung der Zahl der Käufe von 1 auf 2 Käufe einen größeren Effekt auf die Reaktivierungswahrscheinlichkeit hat als die Steigerung von 5 auf 6 bisherige Käufe. Plausibel ist vielmehr, dass eine Verdopplung von 5 auf 10 Käufe denselben Effekt hat wie eine Verdopplung von 1 auf 2 Käufe. Genau diese Annahme wird aber in das Modell gesteckt, wenn eine logarithmierte Einflussgröße verwendet wird. Um dies adäquat umzusetzen, benötigt der Analyst Erfahrung und ein gutes Verständnis der Annahmen des Regressionsmodells.
Dies ist für die alternativen Algorithmen meist nicht notwendig. Hier werden weniger harte Annahmen getroffen, das Modell kann flexibler auf Besonderheiten der zu analysierenden Daten reagieren. Um mit dieser Flexibilität umzugehen, ist allerdings deutlich mehr Rechenzeit und oft eine intensive Gittersuche über die zu setzenden Rahmenparameter erforderlich.
Die logistische Regression als Verfahren arbeitet stabil und birgt i.d.R. kein Risiko der Überanpassung. Neben der kurzen Rechenzeit ist das große Plus der logistischen Regression die gute Interpretierbarkeit. Die aus den Daten ermittelten Regressionsparameter können in relative Veränderungen der Reaktivierungschance bei Änderung der einzelnen Einflussgröße umgerechnet werden. Dies erleichtert Plausibilitätsprüfungen und die Akzeptanz der Analyseergebnisse bei Anwendern.
Bei alternativen Algorithmen muss man hierzu weitere zeitintensive Berechnungen durchführen. Dies ist beispielsweise mit sog. SHAP-Values möglich. Hiermit lassen sich lokale Einflüsse der einzelnen Merkmale darstellen, beispielweise dass der Einfluss einer um einen Kauf angestiegenen Kauf-Anzahl umso geringer ausfällt, je höher die Ausgangsanzahl der Käufe war.
Die bisher von mar,an,con in der Regel verwendete logistische Regression stellt also einen guten Kompromiss zwischen dem Analyseergebnis und dem damit verbundenen Aufwand dar. Natürlich kann man mehr Aufwand investieren, um die Qualität der Scores zu optimieren, vorausgesetzt das beauftragende Unternehmen ist bereit, diesen zusätzlichen Aufwand zu tragen.
Eine Wahrheit bleibt aber in jedem Fall bestehen: Die entscheidendsten Hebel für den Erfolg eines Scorings liegen in der Erarbeitung einer sauberen Datenbasis. Hierbei sind beispielsweise die folgenden Fragen zu klären:

- Welche Merkmale sollen auf welcher Aggregationsebene betrachtet werden?
- Wie sollen fehlende Werte behandelt werden?
- Wie sollen Analyse- und Reaktionszeitraum zeitlich abgegrenzt werden?

Auf diese und ähnliche Fragen müssen bereits geeignete Antworten gefunden werden, lange

bevor der Scoring-Algorithmus ins Spiel kommt. Oder anders ausgedrückt: Wenn die entscheidenden Einflussgrößen in der Analyse fehlen, dann hilft auch ein ausgeklügelter Algorithmus nicht weiter.

Seit 2006 unterstützen wir Unternehmen bei der effizienten Ansprache ihrer Kunden. Unser Service reicht von **Konzeption** und Aufbau einer Kundendatenbank über **Analysedienstleistungen** wie anstoßspezifischen Scorings, Kunden-Segmentierungen oder Kampagnen-Erfolgsmessung bis zur **umfassenden Beratung** und Schulung bei der Umsetzung von Dialogmarketing-Maßnahmen.

Bei der **erfolgreichen Nutzung Ihrer Daten** zur Optimierung Ihres Geschäfts sind sie nicht auf sich allein gestellt. Schon bei der Auswahl geeigneter Fragestellungen helfen wir Ihnen auf Basis der Erfahrung aus zahlreichen Kundenprojekten gern weiter. Wir unterstützen bei der Vorbereitung des Workshops und analysieren im Anschluss daran kompetent Ihre Daten. Die Ergebnisse können wir auch schon während der Analysephase diskutieren und hinsichtlich ihres Nutzens bewerten. Am Ende steht die Implementierung in Ihre Prozesse – auch hierbei erhalten Sie selbstverständlich **kompetente Unterstützung** durch die Experten von **ma,ran,con**.

Ihr Ansprechpartner:

Christian Neumann
Geschäftsführer

Tel. +49 228 338300-50

mar,an,con
Gesellschaft für Marketing, Analyse und Consulting mbH

Königswinterer Straße 418, 53227 Bonn
Tel. +49 228 338300-00
Fax +49 228 338300-99