



Dr. Jan-Paul Lüttke

# Das Random-Forest-Modell: Der Kunde im Meer der Window-Shopper

Wird ein Online-Shopbesucher zum Käufer oder bleibt er ein digitaler Window-Shopper? Mit den richtigen Datenerhebungen und statistischen Verfahren wie dem Random-Forest-Modell können Händler diese Frage beantworten und mit Kampagnen im Shop ressourcenschonend die notwendigen Kaufanreize setzen.

## DER AUTOR



**Dr. Jan-Paul Lüttke** ist einer der Gründer und Geschäftsführer der Aka-noo GmbH, einem schnell wachsenden Digitalunternehmen mit Fokus auf die Bereiche Predictive Analytics und Conversion-Optimierung. Dort verantwortet er die Kundenbetreuung und das Neukundengeschäft sowie die marktgerechte Entwicklung und finanzielle Planung des Unternehmens.

Jeden Tag besuchen Tausende und Abertausende Nutzer Online-Shops. Der Shop der Modemarke Esprit beispielsweise wird laut Wolfram Alpha täglich von 140.000 Besuchern angesteuert, Elektrohandel Conrad verzeichnet 620.000 Visits, Outdoorausstatter Globetrotter 89.000 Visits. Von diesen Besuchern konvertieren in Online-Shops jedoch im Durchschnitt gerade einmal drei Prozent. Aber welcher User in dem Meer der digitalen Window-Shopper wird zum Käufer?

Dies ist eine der sich am häufigsten stellenden Fragen für E-Commerce-Manager. Viele Unternehmen sammeln daher große Mengen an Userdaten. Sie erfassen Klickverhalten, vorherige Einkäufe, Häufigkeit des Besuchs bis zum Kauf, Warenkorbgrößen und bevorzugte Produktkategorien. Geht es an die Analyse, sitzen sie aber oft ratlos vor ihrem Datenschatz. Meist

betrachten sie rückwirkend das Userverhalten auf Unterseiten und optimieren daraufhin, um langfristig Absprungraten zu verringern. Dabei gibt es Möglichkeiten, die Kaufwahrscheinlichkeit, die Neigung zum Warenkorbabbruch oder das Sales-Potenzial eines Users vorauszusagen, während er noch auf der Shop-Seite surft: Prediction-Algorithmen wie das Random-Forest-Modell erlauben das.

## Prediction-Algorithmen im E-Commerce – Anwendung, Chancen und Herausforderungen

Den Begriff Prediction Algorithm liest man derzeit gern und viel, ob es um Attributionsmodelle für die Verteilung von Werbebudgets geht oder um die Identifikation des Käufers unter den Online-Bummeln. Predictive Modelling beschreibt dabei nur die Berechnung von Wahr-

scheinlichkeiten mittels Algorithmen in Echtzeit.

In simplen Fällen werden dazu Formen linearer Prädiktoren wie das lineare Regressionsverfahren eingesetzt. Diese sind in ihren Möglichkeiten jedoch stark begrenzt. So können sie nur eingeschränkt Wechselbeziehungen zwischen Aspekten einfangen, wie sie beispielsweise bei der Kaufentscheidung relevant sind. Hinzu kommt, dass bei linearen Modellen schnell ein sogenannter Overfit eintritt. Overfitting beschreibt eine zu große Übereinstimmung mit einzelnen wahrscheinlichen Ereignissen, die dadurch andere mögliche Ausgänge verzerrt oder vernachlässigt. Ein adäquaterer Ansatz ist das auf dem Zufallsprinzip basierende Random-Forest-Modell.

### Random-Forest-Modell – vom Entscheidungs-Baum zum Wald

Um zu verstehen, wie ein Random Forest funktioniert, gilt es zunächst, Entscheidungsbäume zu verstehen. Ein Entscheidungs- oder Ereignisbaum bildet eine Gruppe von Entscheidungsfragen, sogenannte Variablen, und die sich ergebenden Antwortoptionen schrittweise ab. Von Frage zu Frage verzweigt der Baum sich dabei weiter. Im Falle der Kaufwahrscheinlichkeit können dies Fragen sein wie „Ist der Kunde bekannt?“, gefolgt von „Hat er sich mehr als drei Produkte angeschaut?“ und „Dauert der Besuch länger als fünf Minuten?“ Zum Ende hin lässt sich der User als Käufer oder Nicht-Käufer klassifizieren.

Nun muss das Overfitting, das zuvor beschrieben wurde, umgangen werden. Dazu werden viele dieser Bäume nicht zusammenhängend zu einem sogenannten Random Forest zusammengefügt. Es bedarf dafür einer größeren Anzahl an Datensätzen und Variablen. Ein Beispiel: Ein Händler baut seinen Entscheidungsbaum nur anhand von Daten

#### Daten (Welche Daten sollten erhoben werden?)

Information über vorherige Sessions
Anzahl von Pageviews pro Session
Verschiedene Pagetypes (Produktseite, Übersichtsseite etc.)
Warenkorbveränderungen (Hinzufügen von Produkten, Löschen von Produkten)
Interaktionen wie Klicks, TabSwitches
Metadaten über den Nutzer (z. B. welchen Browser er benutzt)

Tabelle 1: Daten

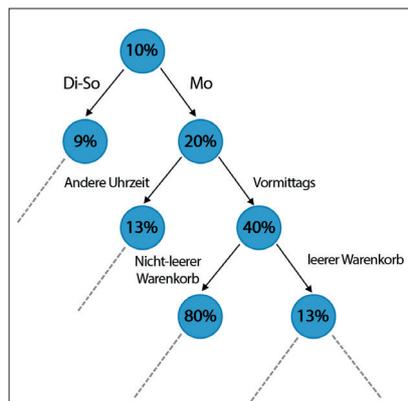


Abb.1: Beispielhafter Entscheidungsbaum anhand von Wochentag, Tageszeit und Warenkorb-Fülle

rund um den 1. Mai auf, der ein Montag ist. Es ist wahrscheinlich, dass die Conversion-Rate und das Kaufverhalten an diesem Montag, einem Feiertag, ganz anders ausfallen werden als in anderen Wochen. Der Händler muss weitere Wochen berücksichtigen, um die Aussagekraft seines Entscheidungsbaums nicht zu gefährden.

Zur Orientierung: Verlässliche Algorithmen aus der Shopanalyse und -optimierung berücksichtigen in ihren zufällig aufgebauten Entscheidungsbäumen über 50 unabhängige Variablen.

### Grundlage Datenerhebung: Welche Daten müssen Shopbetreiber erfassen, um zu erkennen, wer kaufen wird und wer nicht?

Um mit dem Random-Forest-Modell Kaufwahrscheinlichkeiten und andere relevante Fragen beantworten zu können, müssen Händler in drei Schritten denken. Zunächst bedarf es einer Zielsetzung: Einfachere Random-

Forest-Analysen können beispielsweise die Kaufwahrscheinlichkeit, die Wahrscheinlichkeit eines Warenkorbabbruchs oder eines Upsells vorhersagen.

Aus der Zielsetzung ergeben sich die zu erhebenden Datensätze. Im Fall der Kaufwahrscheinlichkeit sollte der Händler die in Tabelle 1 aufgeführten Datentypen für jeden Kunden in seinen bestehenden Systemen tracken.

Die Kombination dieser Datentypen ermöglicht bereits ein aussagekräftiges Bild über die Wahrscheinlichkeit, mit der ein User kauft oder nur online bummt. Darüber hinaus gibt es eine Vielzahl anderer möglicher Datenpunkte, die interessant sein können, herkömmliche Analytics- und Optimierungstools aber überfordern. So gibt es Features aus mehreren Datentypen, die von professionellen Tools zusammengeführt werden. Sie erlauben weitere Schlüsse auf das Kaufverhalten: Die Homogenität von Produktkategorien, der Preis angesehener Produkte im Vergleich zum Durchschnittspreis des Shops oder die Häufigkeit, mit der ein User ein Produkt sessionübergreifend angeschaut hat, sind nur drei Beispiele.

Umsetzung: Mit dem Random-Forest-Modell Käufer und Shopper unterscheiden

Diese Datensätze gilt es nun in Entscheidungsbäumen mit den entsprechenden Variablen zu verknüpfen. Relevante Fragen können sein:

- » Ist der Nutzer zum ersten Mal in dem Shop?
- » Wie viele PageViews hat der Nutzer, seitdem er auf dem Shop ist?
- » Hat der Nutzer bereits etwas in den

Wochentag	Montag	Dienstag	Mittwoch	Donnerstag	Freitag	Samstag	Sonntag
Views, Conversions	1000, 200	1000, 50	1000, 50	1000, 100	1000, 100	2500, 250	2500, 250

Tabelle 2: Beispielhafte Conversion-Wahrscheinlichkeit nach Wochentag

- Warenkorb gelegt?
- » Wie oft war der Nutzer in dem Shop in den letzten zwei Wochen?
- » Wie viele Tab-Switches macht der Nutzer auf einer Produktseite?
- » Wie viele Klicks macht der Nutzer auf einer Produktseite?
- » Welcher Wochentag/welche Uhrzeit ist gerade?
- » Wie viele PageViews hatte der Nutzer, als er das letzte Mal in dem Shop war?
- » Wie schnell klickt der Nutzer sich zu einer Produktseite?
- » Benutzt der Nutzer ein Smartphone, Tablet oder einen PC, um den Shop zu besuchen?

In den Entscheidungsbäumen werden die Fragen in unterschiedlicher Reihenfolge und Zusammensetzung kombiniert. Im Beispielbaum aus Abbildung 1 stellt ein Händler drei Fragen in Verbindung zueinander: An welchem Tag ist der Besucher im Shop? Zu welcher Tageszeit? Und ist sein Warenkorb leer oder voll?

Für jede Frage und ihre Antwortmöglichkeiten enthält der Beispielbaum die Wahrscheinlichkeit der Conversion. Dazu rückt er die Anzahl der Visits und der Conversions an den einzelnen Wochen ins Verhältnis. In Tabelle 2 sind beispielhafte Korrelationen aufgeführt. Für den Händler bietet der Montag demnach die größte Chance auf eine Conversion.

Verfolgt der Händler nun den Entscheidungsweig des Montags weiter, stellt sich die Frage nach der Tageszeit. So ist am Montag der Vormittag am gewinnbringendsten, und zu dieser Zeit – wenig überraschend – ein nicht-leerer Warenkorb.

Auf Basis dieser Daten weiß der Händler nun also, dass jemand, der Montagvormittag mit einem nicht-lee-

Montag	Vormittags	Nachmittags	Abends/Nachts
Views, Conversions	250, 100	250, 50	500, 50

Tabelle 3: Beispielhafte Conversion-Wahrscheinlichkeit nach Tageszeit für den Montag

Montagvormittag	Leerer Warenkorb	Nicht-leerer Warenkorb
Views, Conversions	250, 100	250, 50

Tabelle 4: Beispielhafte Conversion-Wahrscheinlichkeit nach Warenkorbfüllung für den Montagvormittag

ren Warenkorb in seinem Beispielshop unterwegs ist, in 80 von 100 Fällen konvertiert. Auf diese Art und Weise können Shopbetreiber eine Vielzahl von Fragen beantworten und Erkenntnisse für die weitere Arbeit gewinnen.

Kaufwahrscheinlichkeit verstanden. Und nun?

So können Shop-Betreiber ihre Energie und Ressourcen auf Personen fokussieren, die bei einem der nächsten Besuche mithilfe von Kampagnen konvertieren. Das gilt insbesondere für Traffic einkaufende Retargeting- oder Affiliate-Maßnahmen, die per Klick oder Conversion abgerechnet werden.

Noch spannender wird es, wenn der Random Forest nicht die blanke Kaufwahrscheinlichkeit eines Users errechnet, sondern vielmehr, mit welcher Wahrscheinlichkeit dieser Besucher einen Anreiz braucht und welche Kampagne die richtige ist, um ihn zur Conversion zu führen. Die Ergebnisse können in Echtzeit in den Besuch des einzelnen Users einfließen, indem Top-Seller beworben, ein zum Kaufentschluss notwendiger Rabattcoupon angeboten oder zum Besucherprofil besser passende Produkte vorgestellt werden.

Mittelfristig erlauben Prediction-Algorithmen in Kombination mit Identifikatoren wie Cookies oder User-IDs ein umfangreicheres Profiling des Shopbesuchers. Ist der User jemand, der nur bei Promotions aktiv wird? Klickt er

sich regelmäßig durch den Shop? Wie markentreu ist er? Und wie viel Umsatz bringt er im Vergleich zu anderen Käufern?

Auch ein Verständnis der wichtigsten Mechanismen und Pfade im eigenen Shop erlauben Prädiktoren. So kann der Händler sich in der Shopoptimierung auf die Seiten und Stellschrauben konzentrieren, die besonders kaufrelevant sind, statt nur auf die, die den meisten Traffic bei sich bündeln.

### Fazit

Jeder Shopbetreiber kann Kaufwahrscheinlichkeiten erkennen und für die Führung der Customer Journey sowie den Traffic-Einkauf effizienter nutzen, wenn er das Random-Forest-Modell einsetzt. Entscheidend für aussagekräftige Prognosen sind immer die konkrete Zielsetzung, die Datenqualität und -menge und die Einbindung in eine entsprechende Anzahl miteinander unverbundener Entscheidungsbäume. Damit lassen sich für Online-Shops Käufer von Window-Shoppern unterscheiden und der Gesamtumsatz im Shop gezielt steigern. ¶