

Machine Learning zur Mailing-Selektion

Datenschutzkonform arbeiten mit Algorithmus aus bestehenden Fundraisingdaten

Spendenbriefe bleiben unverzichtbar. Ein schlankes Machine-Learning-Modell zeigt, wie NGOs ihre Mailing-Auflagen senken und/oder zugleich den Überschuss steigern können – ganz ohne Daten in die Cloud zu schieben und mit voller Transparenz für das Fundraising-Team. Die Praxisprobe zeigt: minus 18,9 Prozent Briefe, plus 6,3 Prozent Überschuss.

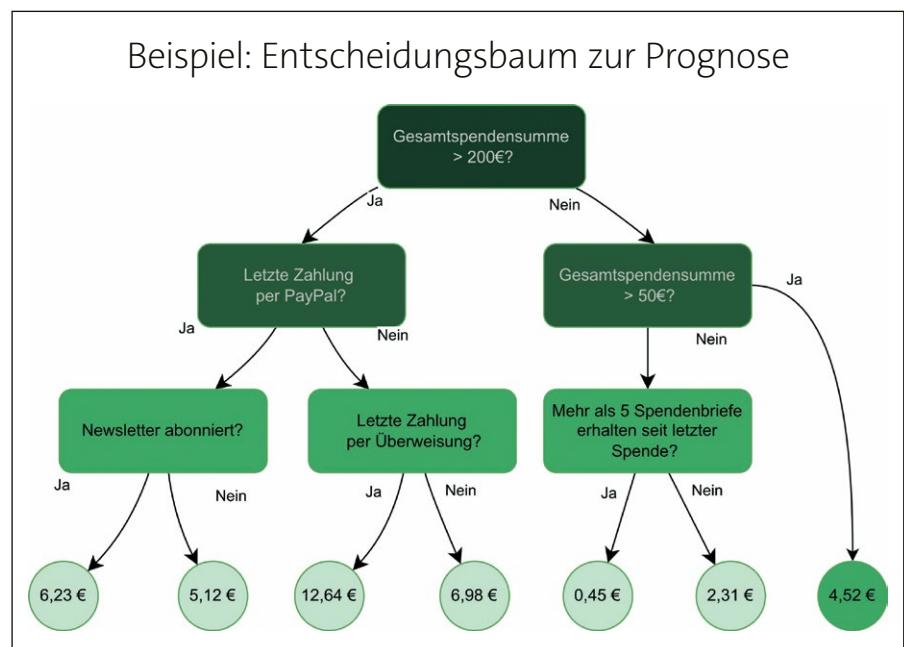
Von OLIVER MOLDENHAUER und DAN WILSON

Fundraising-Briefe bleiben einer der wirksamsten Kanäle im Non-Profit-Sektor, gleichzeitig verursachen sie erhebliche Porto- und Druckkosten sowie CO₂-Emissionen. Die Lösung: nur jene Menschen anschreiben, bei denen statistisch ein hoher Spendenertrag zu erwarten ist. Viele Organisationen verlassen sich dafür noch auf simple Regeln wie „Wer fünf Mailings ignoriert hat...“. Diese Heuristiken bleiben oberflächlich, verschenken Potenzial – und treiben die Kosten.

Selbst datensparsame NGOs besitzen reichhaltige Informationen: Postleitzahl, Mailing-Historie, Spendensummen nach Zahlungsweg, Mitgliedsstatus, Zeitpunkt der letzten Spende usw. Statt diese Daten per Bauchgefühl zu filtern, lässt sich daraus ein Prognose-Modell trainieren, das für jede Person den zu erwartenden Erlös pro Brief berechnet – statistisch fundiert und selbstlernend.

Eigenes Machine Learning statt ChatGPT & Co.

Eine Fundraising-Datenbank einfach in ChatGPT hochzuladen, verbietet sich: Es handelt sich um sensible personenbezogene Daten, häufig auch nicht im von ChatGPT „verstandenen“ Format. Dazu kommt die Blackbox-Problematik – ein Sprachmodell verrät nicht, wie es zu seinen Empfehlungen



kommt. Unser Ansatz: Wir nutzen Methoden des Machine Learning. Hierbei wird mit einem (relativ) einfachen Algorithmus aus den Fundraisingdaten ein mathematisches Modell entwickelt, das für jede Person den erwarteten Ertrag prognostiziert. Die Daten und die Berechnungen landen zu keiner Zeit in der Cloud eines US-Internetriesen, wo sie dem Zugriff der amerikanischen Behörden ausgeliefert wären. Wir arbeiten mit Open-Source-Software, die auf einem verschlüsselten Linux-PC lokal in unserem Büro läuft.

Praxisbeispiel

Wir können dabei viele verschiedene Methoden des Machine Learnings anwenden. Typischerweise gibt es einen gewissen Trade-Off zwischen sehr leicht erklärbaren Modellen, wie Entscheidungsbäumen, die gute Ergebnisse liefern, und weniger leicht erklärbaren, die bessere Ergebnisse liefern, aber schwieriger zu erklären sind. Welche Methode am besten passt, hängt von der NGO und ihren Daten ab.

In unserem Beispiel arbeiten wir für eine mittelgroße NGO, deren Einnahmen überwiegend aus Spenden und Beiträgen stammen, viele davon ausgelöst durch Spendenbriefe. Datenschutz und Transparenz gehören zu den Grundwerten unserer Auftraggeberin. Für die neue Mailing-Selektion galten daher zwei Bedingungen: keine personenbezogenen Daten in der Cloud. Und: vollständige Nachvollziehbarkeit der Entscheidungen.

Wir setzen also auf den einfachsten Modelltyp: den Entscheidungsbaum. Praktisch ist das eine Sammlung von Wenn-Dann-Regeln. Die Grafik oben zeigt ein fiktives Beispiel mit drei Fragen und sieben Endpunkten. (Das tatsächlich genutzte Modell hat bis zu sieben Fragen und 82 Endpunkten) Lesehilfe: Wenn eine Person bisher schon insgesamt über 200 Euro gespendet hat, die letzte Zahlung per Paypal war und sie den Newsletter nicht abonniert hat, dann wird eine Spende von 5,12 Euro pro Brief erwartet.

Wir programmierten eine Schnittstelle zur Fundraising-Datenbank der NGO und lasen

die wesentlichen Zuwendungsdaten der letzten Jahre aus. Daraus berechneten wir ein Entscheidungsbaum-Modell – zunächst nur für Personen, die bereits mindestens einmal gespendet hatten. Für Personen ohne Spendenhistorie lagen bislang zu wenige Merkmale vor, um verlässlich zu prognostizieren; wir entwickeln derzeit ein ergänzendes Modell, das künftig auch diese Zielgruppe abdeckt. Vor jedem Mailing definiert die Organisation den statistischen Mindestertrag, ab dem ein Brief verschickt wird. Regelmäßig wird das Modell mit den neu eingegangenen Daten aktualisiert.

**Ergebnisse: 6,3 Prozent mehr Wirkung,
18,9 Prozent weniger CO₂**

Bei den ersten zwei Mailings mit dem neuen Modell wurden 23,7 Prozent der bisherig angeschriebenen Personen aus dem Verteiler genommen – ihr Durchschnittserlös lag

zuvor bei nur 0,01 Euro pro Brief. Gleichzeitig wählte das Modell 4,8 Prozent neue Personen aus, die im Mittel das 3,7-fache des Schwellenwerts spendeten. Unterm Strich sank das Mailing Volumen um 18,9 Prozent, und der Netto-Überschuss stieg um 6,3 Prozent.

Ein datengetriebenes Modell spart also Porto, reduziert CO₂ und steigert die Erlöse. Es lernt regelmäßig hinzu und lässt sich flexibel an Mindestwerte oder Vorgaben zur Zahl der versendeten Briefe anpassen. Gleichzeitig bleiben Nachteile vieler KI-Anwendungen außen vor: Daten verbleiben sicher vor Ort, der Energieverbrauch bleibt übersichtlich, und der Entscheidungsbaum bietet volle Transparenz statt Blackbox.

Als nächste Schritte entwickeln wir einen Selektions-Prototypen für Wikimedia Deutschland und erweitern das bestehende Modell um eine Selektion für potenzielle Neumitglieder. □



Oliver Moldenhauer ist Gründer und Geschäftsführer von DataStrategies4Change. Ursprünglich Physiker, arbeitet er seit über 35 Jahren mit zivilgesellschaftlichen Organisationen. Er war u.a. Mitgründer von Attac Deutschland, WeMove Europe und Investigate Europe sowie Campagnier bei Campact und Ärzte ohne Grenzen. Heute kombiniert er tiefes NGO-Know-how mit Data Science und verantwortet Strategie, Beratung und Modellentwicklung bei DS4C.
► www.datastrategies4change.eu



Dan Wilson ist Physiker und Neurowissenschaftler mit über zehn Jahren Erfahrung in Forschung, Programmierung und Data Science. Sein Schwerpunkt liegt auf Machine Learning und erklärbarer KI, die er bei DS4C in Spendenvorhersagen und Visualisierungen einsetzt.
► www.datastrategies4change.eu