

Revenue Operations & KI

Was ist Lead-Scoring?

Lead-Scoring ordnet Leads nach der Chance auf ein benanntes Ergebnis. Ohne dieses Ergebnis, ein Zeitfenster und eine Grundgesamtheit sagt ein Score wenig.

Autor	Veröffentlicht	Lesezeit	Wörter
Sinan Isoglu, MBA (Quantic)	18. Oktober 2026	19 Min.	4.044
Führungskraft für Umsatzwachstum, Dozent und Doktorand			

Management Summary

Lead-Scoring ordnet jedem Lead eine Schätzung zu, wie wahrscheinlich er ein benanntes Ergebnis erreicht oder was es wert wäre, damit sich Leads ordnen lassen. Dieser Konzeptleitfaden definiert den Begriff, trennt Scores nach Regeln von Scores aus Modellen, die aus Daten geschätzt werden, und grenzt Scoring von Qualifizierung und Routing ab. Er berichtet über eine Befragung in vier B2B-Unternehmen, in der Vertriebsmitarbeitende, die die Lead-Vorauswahl besser bewerteten, Marketing-Leads einen größeren Teil ihrer Zeit widmeten, je Unternehmen unterschiedlich stark, und zwei Modelle, deren Genauigkeit erst im Vergleich mit dem häufigsten Ergebnis etwas aussagt. Dazu kommt ein Blatt mit den Angaben, die ein Score braucht, bevor jemand danach handelt. Die Quellen sind eine Befragung, ein an simulierten und an realen Anbieterdaten getestetes Modell und die CRM-Daten eines Unternehmens.

Inhalt	§ 1	Was ist Lead-Scoring?	3
	§ 2	Was muss ein Score benennen, bevor er etwas bedeutet?	4
	§ 3	Wer legt die Gewichte fest: Menschen oder Daten?	6
	§ 4	Ist ein Score dasselbe wie Qualifizierung oder Routing?	7
	§ 5	Was geschieht mit einem Lead nach dem Score?	7
	Abbildung 1	Wie stark der Zeitanteil für Marketing-Leads mit der Bewertung der Vorauswahl stieg	9
	§ 6	Warum kann ein Score treffsicher wirken und trotzdem in die Irre führen?	11
	§ 7	Welche Angaben gehören zu einem Score, bevor er an den Vertrieb geht?	13
	Abbildung 2	Ein Blatt für einen Lead-Score	14

§ 8	Was müsste ein Softwareteam noch prüfen?	15
	Referenzen	16

Was ist Lead-Scoring?

§ 1 von 8

Lead-Scoring ordnet jedem Lead eine Zahl oder eine Klasse zu, die schätzt, wie wahrscheinlich er ein benanntes Ergebnis erreicht, etwa ein erstes Gespräch oder einen unterschriebenen Auftrag, oder was dieses Ergebnis wert wäre. Mit der Schätzung werden Leads in eine Reihenfolge gebracht, damit die Menschen, die sie nachverfolgen, dort anfangen, wo sich der Aufwand am ehesten lohnt. Keine der drei Studien, auf die sich dieser Leitfaden stützt, verwendet den Begriff. Jede behandelt einen Teil davon: wie Vertriebsmitarbeitende auf die Vorauswahl reagieren, die bestimmt, welche Leads bei ihnen ankommen, ein Modell, das die Kaufphase eines Interessenten aus seinem Verhalten schätzt, und ein Modell, das erklärt, welche Leads gewonnen, verloren oder abgebrochen wurden.

Der Schritt der Vorauswahl, den Sabnis et al. (2013)² untersuchten, ist nach der Definition dieses Leitfadens die Stelle, an der ein Score auf seinen Schwellenwert trifft. Vorqualifizierung ist bei ihnen ein Prozess, in dem die Marketingabteilung die von ihr erzeugten Leads sichtet und jene aussortiert, die sie für unattraktiv hält („Prequalification is a process by which the marketing department screens the leads it has generated and discards those it deems unattractive.“). Sie findet im Unternehmen selbst statt, in der Regel durch Marketingpersonal („lead prequalification is typically executed internally by marketing personnel“). Die Befragungsisems der Autoren beschreiben, was eine solche Vorauswahl aus Sicht des Vertriebs leisten soll, etwa, dass das eigene Unternehmen sehr gut darin sei, dem Vertrieb nur heiße Leads zu geben („My company is very good at giving only hot leads to salespeople.“).

² Sabnis, G., Chatterjee, S. C., Grewal, R., & Lilien, G. L. (2013). The sales lead black hole: On sales reps' follow-up of marketing leads. *Journal of Marketing*, 77(1), 52–67.

Marvasti et al. (2021)¹ stellen die Scoring-Frage schon im Titel („Is this company a lead customer?“) und beantworten sie mit einem Modell: einem Verfahren des maschinellen Lernens, das die Phasen der B2B-Kaufreise aus dem Surfverhalten der kaufenden Unternehmen erschließt („the authors propose a machine learning approach to infer the stages of the B2B buying journey by observing the online browsing behavior of buyer companies“). Die Phase selbst ist nicht beobachtbar, sie bleibt dem Anbieter verborgen („These buying stages are hidden from the sellers“); das Surfverhalten steht stellvertretend für sie. Virtanen et al. (2015)³ modellierten das Ergebnis jedes einzelnen von 4.194 Vertriebsleads eines IT-Lösungsanbieters („The final dataset

contains 4194 sales leads, and each lead is nested under a certain customer.“). Ein solches Modell, gebaut auf früheren Leads und ihren Ergebnissen, ist das, worauf ein modellbasierter Score beruht.

¹ Marvasti, N. B., Huhtala, J.-P., Yousefi, Z. R., Vaniala, I., Upreti, B., Malo, P., Kaski, S., & Tikkanen, H. (2021). Is this company a lead customer? Estimating stages of B2B buying journey. *Industrial Marketing Management*, 97, 126–133.

³ Virtanen, T., Parvinen, P., & Rollins, M. (2015). Complexity of sales situation and sales lead performance: An empirical study in business-to-business company. *Industrial Marketing Management*, 45, 49–58.

In diesem Leitfaden ist ein Lead-Score eine Schätzung für einen einzelnen Lead: wie wahrscheinlich er innerhalb eines festgelegten Zeitfensters ein benanntes Ergebnis erreicht oder was dieses Ergebnis wert wäre, erstellt für eine festgelegte Grundgesamtheit von Leads und aus Informationen, die zum Zeitpunkt der Bewertung vorliegen. Lead-Scoring ist die Praxis, solche Schätzungen zu erstellen und Leads danach zu ordnen. Ein Score ordnet, er entscheidet nicht. Zur Entscheidung wird er erst durch einen Schwellenwert, das ist Qualifizierung, und durch eine Zuständigkeit, das ist Routing.

Wahrscheinlichkeit und Wert sind verschiedene Schätzungen, und nicht alle Vertriebsmitarbeitenden gewichten sie gleich. Sabnis et al.² halten unter Verweis auf frühere Forschung fest, dass Vertriebsmitarbeitende mit besseren früheren Verkaufsergebnissen eher Leads nachgehen, die seltener zum Abschluss führen, aber mehr Umsatz versprechen, als umgekehrt („Sales reps with better previous sales performance are more likely to follow up on leads that offer a lower likelihood of closure but higher revenue potential, rather than the reverse“). Nach der Lesart dieses Leitfadens kann ein Score, der nur nach Wahrscheinlichkeit ordnet, Leads deshalb anders reihen, als es die stärksten Vertriebsleute eines Unternehmens täten.

Was muss ein Score benennen, bevor er etwas bedeutet?

§ 2 von 8

Drei Dinge: das Ergebnis, das er vorhersagt, das Zeitfenster, in dem es eintreten muss, und die Grundgesamtheit der Leads, für die er gebaut wurde. Die Quellen zeigen, wie leicht sich jedes davon verschiebt.

Zuerst das Ergebnis. In den realen Daten von Marvasti et al.¹ stammten die vorherzusagenden Phasen aus den Opportunity-Datensätzen eines Anbieters, und ein Vertriebsspezialist half, 15 Opportunity-Typen den vier Kaufphasen zuzuordnen („there are 15 different opportunity types defined in this dataset mapped to previously described, four buying

states with the help of a sales specialist“). In dieser Zuordnung zählen „Closed Won“ und „Closed Lost“ beide als „Late Funnel“ (Tabelle 6, S. 131). Ein Modell, das die späte Phase richtig erkennt, erkennt damit noch nicht, wer kauft. Die Befragung von Sabnis et al.² legte das Ende der Nachverfolgung mit drei Ausgängen fest: Sie beginnt mit dem ersten Kontakt und endet, wenn feststeht, was der Interessent tun wird, also wenn man den Auftrag erhält, ein anderer ihn erhält oder man den Interessenten als aussichtslos abschreibt („We count follow-up as starting when you contact a prospect in any way at all and ending when you know what the prospect will do (you get the order, somebody else gets the order, or you write them off as hopeless).“). Ein Score, der den ersten dieser Ausgänge vorhersagt, ist ein anderer Score als einer, der nur vorhersagt, dass ein Lead nicht abgeschrieben wird.

Dann das Zeitfenster. Marvasti et al.¹ prüften die Schätzung jedes Tages an der Phase einer Opportunity, deren Datum höchstens zwei Wochen davor oder danach lag („Note that we have considered an interval of 2 weeks before and after each opportunity date as a valid interval for each buying stage.“). Mit einem anderen Intervall stünde dasselbe Modell vor einem anderen Test; das Zeitfenster gehört deshalb neben jede Angabe zur Genauigkeit.

Schließlich die Grundgesamtheit. Ein Modell, das auf Surfverhalten beruht, kann nur Unternehmen bewerten, die surfen: In den realen Daten von Marvasti et al.¹ waren das 648 Unternehmen, die in gut einem Jahr die Website des Anbieters besuchten. Virtanen et al.³ grenzten ihre Leads vor der Modellierung ein und schlossen Leads mit fehlenden Angaben oder geringem Umsatz aus („Leads that did not include all necessary data points, or had low revenue, were ruled out of the study.“). Ein Score gilt für die Art von Leads, auf der er gebaut wurde, und sagt über die ausgeschlossenen zunächst nichts.

Wer legt die Gewichte fest: Menschen oder Daten?

§ 3 von 8

Bei einem regelbasierten Score legen Menschen fest, was zählt und wie viel: Punkte für eine Position, eine Unternehmensgröße oder einen Download und eine Schwelle, ab der das Marketing den Lead weitergibt. Die Vorqualifizierung bei Sabnis et al.² ist in diesem Sinne eine von Menschen gesetzte Vorauswahl, betrieben von Marketingpersonal, und die befragten Vertriebsmitarbeitenden beurteilten sie danach, ob sie filtert: Ein Item lautete, das eigene Unternehmen sortiere kalte Leads miserabel aus („My company does a terrible job of filtering out cold leads“), und wurde umgekehrt gewertet. Bei einem modellbasierten Score stammen die Gewichte aus Daten. Im Hidden-Markov-Modell von Marvasti et al.¹ werden die Parameter nicht vorab festgelegt, sondern aus den beobachteten Daten selbst geschätzt („it should initially be estimated from the data (observed buying behavior) itself“). Virtanen et al.³ erklärten das erfasste Ergebnis jedes Leads mit einer multinomialen logistischen Regression.

Beide brauchen Verschiedenes. Eine Regel braucht jemanden, der festlegt, was zählt, und bleibt eine Vermutung, bis sie an Ergebnissen geprüft ist. Ein Modell braucht Daten, aus denen es lernt, und Ergebnisse, an denen es geprüft wird, und auch die Ergebnisse erfassen Menschen. Das Modell von Marvasti et al.¹ lernte allein aus dem Surfverhalten; die Opportunity-Datensätze des Anbieters dienten nur zur Prüfung, und die Autoren warnen, dass die Vergleichsdaten ungenau sein können und der Anbieter an manchen Tagen bestehende Opportunities übersehen haben kann („the provided ground truth may not be precise, and the seller may have missed existing opportunities on some dates“). Eine Genauigkeit, die an lückenhaften Daten gemessen wird, misst die Datensätze ebenso wie das Modell. Ein Modell, das an solchen Daten geschätzt wird, lernt die Datensätze ebenso wie die Käufer; so wurde das Modell von Virtanen et al.³ an den Ergebnissen geschätzt, die Vertriebsleute erfasst hatten. Keine der Quellen vergleicht einen regelbasierten mit einem modellbasierten Score an denselben Leads; dieser Leitfaden sagt deshalb nicht, welcher von beiden besser ordnet.

Ist ein Score dasselbe wie Qualifizierung oder Routing?

§ 4 von 8

Nein. Scoring schätzt, Qualifizierung entscheidet, Routing weist zu. Qualifizierung ist die Ja-oder-Nein-Entscheidung, dass ein Lead gut genug ist, um weitergegeben zu werden; Sabnis et al.² nennen das Vorqualifizierung. Ein Score kann über einen Schwellenwert in sie einfließen, aber der Schwellenwert ist eine eigene Entscheidung mit eigenen Kosten: Jeder Lead knapp darunter gilt als keinen Anruf wert. Routing gibt einem zulässigen Lead dann eine Zuständigkeit oder eine Queue. Der Leitfaden zum Lead-Routing zieht dieselbe Grenze von der anderen Seite: Ein Score kann Datensätze ordnen oder priorisieren, Routing weist sie zu, und der Zuweisungsdatensatz zeigt nicht, ob ein Lead gut war. Der Leitfaden zum Product-Qualified Lead behandelt eine Art von Qualifikationsregel, eine Schwelle für die Produktnutzung in einem festgelegten Zeitfenster.

Was geschieht mit einem Lead nach dem Score?

§ 5 von 8

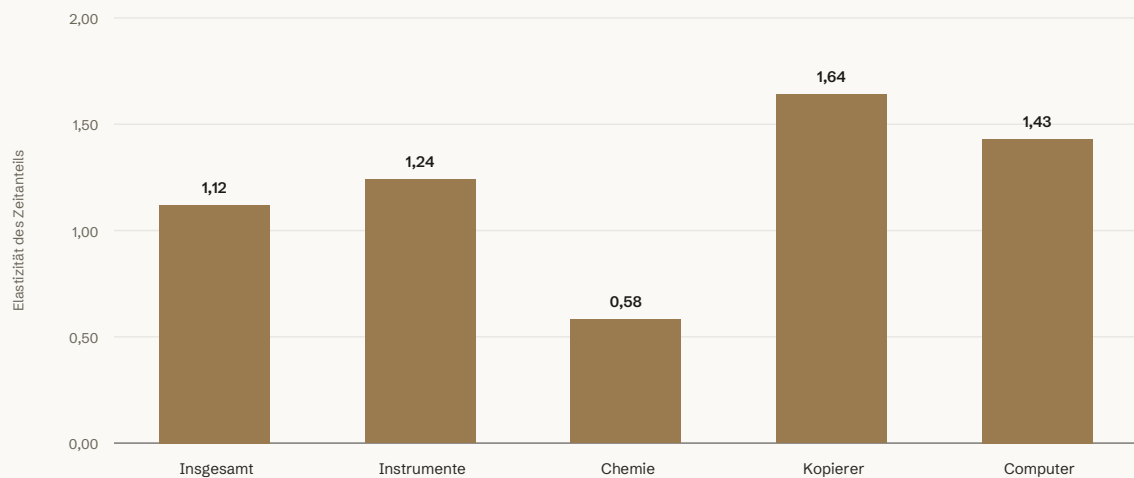
Er landet bei einem Menschen, der entscheidet, ob er Zeit darauf verwendet, und der ihn auch danach beurteilt, was er von der Vorauswahl hält, durch die der Lead kam. Sabnis et al.² befragten Vertriebsmitarbeitende in vier B2B-Unternehmen, je eines aus den Branchen wissenschaftliche Instrumente, Chemie, Kopierer und Computer, und werteten die Antworten von 461 von ihnen aus. Ihr Ausgangspunkt: Vertriebsmitarbeitende können die objektive Qualität von Leads nicht beobachten; sie richten sich nach ihrer Wahrnehmung der Lead-Qualität oder, praktischer gefasst, des Qualifizierungsprozesses im Marketing („Sales reps cannot observe the objective quality of leads; they are motivated by their perceptions of lead quality or, more operationally, of the marketing department’s lead qualification process.“). Diese Wahrnehmung erhoben sie mit Befragungssitems und setzten sie in Beziehung zum Anteil der Arbeitszeit, den jede befragte Person Marketing-Leads widmete, im Vergleich zu selbst gewonnenen Leads und aller übrigen Arbeit.

Wer die Vorauswahl besser bewertete, verbrachte einen größeren Teil der Zeit mit Marketing-Leads; die Autoren werten das als Bestätigung ihrer ersten Hypothese („which suggests that as the perceived quality of prequalification increases, the proportion of time spent on marketing leads increases“). Als Elastizität für die gesamte Stichprobe ausgedrückt, geht eine um 1 Prozent höhere wahrgenommene Qualität der Vorqualifizierung mit einem um 1,12 Prozent größeren Zeitanteil für Marketing-Leads einher („a 1% increase in perceived prequalification

quality produces a 1.12% increase in marketing lead follow-up time allocation“), und die Elastizität ist in allen vier Unternehmen positiv, schwankt aber deutlich, von 0,58 bis 1,64 („The elasticity is positive across all four firms but varies considerably (from .58 to 1.64).“). Die Autoren schreiben „produces“; die Daten stammen aus einer einzigen Befragung zu einem Zeitpunkt, deshalb liest dieser Leitfaden die Zahl als Zusammenhang. Bei erfahreneren Vertriebsmitarbeitenden war der Zusammenhang stärker: Die Autoren fanden Bestätigung für ihre Annahme, dass Erfahrung die positive Wirkung der wahrgenommenen Qualität der Vorqualifizierung auf den Zeitanteil für Marketing-Leads verstärkt („experience enhances the positive effect of the perceived quality of prequalification on the proportion of time spent following up marketing leads“). Abbildung 1 zeigt die Elastizität insgesamt und je Unternehmen.

Abbildung 1 Wie stark der Zeitanteil für Marketing-Leads mit der Bewertung der Vorauswahl stieg

In jedem Unternehmen widmeten Vertriebsmitarbeitende, die die Vorauswahl besser bewerteten, Marketing-Leads einen größeren Teil ihrer Zeit, doch nach den Punktschätzungen der Autoren unterschied sich die Stärke dieses Zusammenhangs zwischen den Unternehmen etwa um das Dreifache.



Eine Elastizität von 1,12 bedeutet, dass im Modell der Autoren eine um 1 Prozent bessere Bewertung der Vorauswahl mit einem um 1,12 Prozent größeren Zeitanteil für Marketing-Leads einherging. Die Bewertung ist die eigene Antwort jeder befragten Person auf Befragungssitems, kein Maß für die Leads. Dieselbe Tabelle weist über die gesamte Stichprobe für die Kontrolle durch Führungskräfte eine Elastizität von $-0,14$ und für die Lead-Menge eine von $-0,07$ aus. Tabelle 3 nennt für keine Elastizität ein Intervall oder einen Test. Die Werte je Unternehmen stammen aus einem Modell, das jedes Unternehmen zur Schätzung für die gesamte Stichprobe hinzieht, Unternehmen 2 beruht auf 30 Befragten, und der Gesamtwert ist eine eigene Schätzung des Modells, kein Durchschnitt der vier Unternehmenswerte.

Quelle Sabnis et al. (2013), S. 63, Tabelle 3, Zeile Quality of prequalification: Elastizität des Anteils der Arbeitszeit für Marketing-Leads in Bezug auf die wahrgenommene Qualität der Lead-Vorqualifizierung des eigenen Unternehmens, im Dirichlet-Regressionsmodell der Autoren auf Grundlage von Befragungsdaten. Ausgewertet: 51 Vertriebsmitarbeitende in Unternehmen 1 (wissenschaftliche Instrumente), 30 in Unternehmen 2 (Chemie), 100 in Unternehmen 3 (Kopierer) und 280 in Unternehmen 4 (Computer), insgesamt 461 (Tabelle 1, S. 59; Branchen S. 59 und 63).

Zwei Grenzen verhindern, dass daraus ein Befund über Scores wird. Die Studie maß Zeitanteile, nicht Leads: Die Autoren schreiben, ihre Folgerungen ließen sich nicht ohne Weiteres auf die tatsächlich für Marketing-Leads aufgewendeten Stunden oder die Zahl nachverfolgter Leads übertragen („may not extend fully to related constructs such as actual time spent on marketing leads (in hours) or the number of leads

followed“), und erst Daten auf Lead-Ebene aus den CRM-Systemen der Unternehmen könnten solche Analysen ermöglichen („Lead-level data from firms’ customer relationship management systems might make such analyses possible.“). Und die Vorauswahl wurde von den Vertriebsmitarbeitenden bewertet, nicht an Ergebnissen geprüft: Die Qualität der Vorqualifizierung ist kein objektives Maß auf Unternehmensebene, sondern beruht auf den Wahrnehmungen einzelner Vertriebsmitarbeitender („the extent of prequalification quality is not an objective, firm-level measure but rather is based on individual sales reps’ perceptions“), und sie zählen solche Maße zu ihren Grenzen, weil sie mit Messfehlern und womöglich psychologischen Verzerrungen behaftet sind („inherently suffer from measurement error and perhaps some psychological bias“). Was die Studie zeigt: Der Zeitanteil, den Vertriebsmitarbeitende Marketing-Leads widmen, hängt damit zusammen, wie gut sie die Vorauswahl hinter diesen Leads bewerten, und zwar in jedem Unternehmen unterschiedlich stark. Die beiden anderen Hebel der Studie, die Kontrolle der Nachverfolgung durch Führungskräfte und die Zahl der gelieferten Leads, behandelt der Beitrag zum Funnel-Engpass, den niemand misst.

Wer gar keine Zeit für Marketing-Leads angab, kommt in diesen Zahlen nicht vor. 39 Befragte gaben an, keine Zeit mit Marketing-Leads zu verbringen („Our preliminary examination of the data showed that 39 respondents reported spending no time pursuing marketing leads.“). Sie bewerteten die Vorqualifizierung ihres Unternehmens signifikant schlechter als die übrigen, und 37 der 39 kamen aus einem einzigen Unternehmen, dem aus der Computerbranche („37 of the 39 respondents who indicated no marketing lead follow-up time came from one firm (computer industry)“); die Autoren sahen darin die Möglichkeit einer systematischen Verzerrung („a potential for systematic bias“) und nahmen diese Antworten aus der Analyse. Der Ausdruck im Titel des Aufsatzes, das schwarze Loch der Vertriebsleads („sales lead black hole“), bezeichnet die Leads, die das Marketing erzeugt und der Vertrieb nie verfolgt. Den Anteil, den der Aufsatz dafür nennt, entnimmt er zwei zitierten Berichten von Marktforschungsunternehmen, nicht den eigenen Befragten, und die Autoren fügen hinzu, dass die Zahl Schwächen des Marketings ebenso wie des Vertriebs widerspiegelt („reflects inadequacies of marketing as well as the sales function“), weil Vertriebsmitarbeitende so viele Leads erhalten können, dass sie nicht alle verfolgen könnten, selbst wenn sie wollten („it is not possible for the sales reps to pursue all of them, even if they desired to do so“).

Marvasti et al.¹ erwarten, dass Anbieter mit geschätzten Kaufphasen die Trefferquote von Marketing und Vertrieb verbessern können („improve the hit rate of marketing and sales activities“), und für den Zusammenhang zwischen einem auf die Kaufphase abgestimmten Marketing und einer höheren Trefferquote verweisen sie auf frühere Arbeiten. Ihre Studie schätzt Phasen; sie beobachtet nicht, was Anbieter mit den Schätzungen taten, und keine der drei Quellen prüft, ob das Handeln nach einem Score verändert, wie viele Leads konvertieren.

Warum kann ein Score treffsicher wirken und trotzdem in die Irre führen?

§ 6 von 8

Weil die Genauigkeit davon abhängt, wie viel der Score sieht, woran er geprüft wird und wie häufig das Ergebnis ist. Marvasti et al.¹ testeten ihr Modell zuerst an simulierten Käufern, deren wahre Phase bekannt ist; jeder simulierte Datensatz umfasste 300 Unternehmen. Die Genauigkeit lag bei 83 Prozent, wenn das Surfverhalten täglich beobachtet wurde, und bei 71 Prozent, wenn es an etwa einem Drittel der Tage beobachtet wurde; die Autoren fassen zusammen, dass die Leistung umso geringer ausfällt, je weniger Surfaktivität beobachtet wird („the fewer browsing activities are observed, the lower the performance is achieved“). In den Simulationen wurden die Parameter auf demselben Datensatz gelernt, auf dem auch die Phasen geschätzt wurden („Both parameter learning and buying state estimations are performed on the same dataset“); das Modell wurde also an denselben Unternehmen geprüft, aus deren Surfverhalten es gelernt hatte, allerdings ohne ihre wahren Phasen zu kennen. In den realen Daten eines Anbieters, in denen pro Monat an höchstens vier Tagen überhaupt Surfverhalten auftauchte („in each month, at most, four days include some browsing behavior“), lag die berichtete Genauigkeit bei 62 Prozent.

Ein Anteil richtiger Antworten sagt wenig, bevor er mit der Basisrate verglichen wird: wie häufig jedes Ergebnis vorkommt und damit, wie oft eine Regel richtig läge, die stets das häufigste nennt. Virtanen et al. stellten für ihr eigenes Modell zwei solche Vergleiche an. Es ordnete 63,0 Prozent der Leads richtig zu, 50 Prozent besser als der Zufall („which is 50% better than by chance alone“), wobei Zufall die 41,6 Prozent bedeutete, die ein Raten im Verhältnis der Anteile der drei Ergebnisse träfe. Gegenüber dem häufigsten Ergebnis, einer Regel, die jeden Lead als abgebrochen vorhersagt („which predicts that every single lead is canceled“), fiel der Vorsprung kleiner aus: Das Modell verfehlte die Verbesserung um 25 Prozent, die die Autoren als Maßstab

ansetzten, und kam auf 17 Prozent („Our model failed to meet a 25% improvement over this test model, as the improvement rate of our model was 17%.“). Marvasti et al.¹ berichten keinen solchen Vergleich. Die Rechnung dieses Leitfadens an ihrer Konfusionsmatrix für die realen Daten (Tabelle 7, S. 132) findet die meisten bewerteten Tage in einer Phase, Late Funnel: Eine Regel, die an jedem Tag diese Phase nennt, hätte an 65 bis 76 Prozent der Tage richtig gelegen, das Modell an 62 Prozent. Die Spanne ergibt sich aus der Tabelle selbst. Nach ihren Beschriftungen sind die Spalten die tatsächlichen Phasen, doch dann enthält ihre erste Spalte 342 Tage der Phase No-funnel, die laut den Autoren in den Opportunity-Daten gar nicht erfasst ist; die Tabelle lässt sich also in beide Richtungen lesen.

Beide Modelle ordnen jeden Fall einer von mehreren Klassen zu. Ein Lead-Score kann Leads stattdessen nach ihrer Chance auf ein einzelnes seltenes Ergebnis ordnen, etwa einen bezahlten Vertrag, und dann ist das häufigste Ergebnis, dass nichts geschieht: Eine Regel, die stets dieses Ergebnis vorhersagt, läge meistens richtig und würde doch keinen einzigen Lead hervorheben. Nach der Lesart dieses Leitfadens fragt der Vergleich mit der Basisrate bei einem solchen Score deshalb, ob die Leads, die er am höchsten einstuft, unter Leads, an die er nicht angepasst wurde, das Ergebnis häufiger erreichen als diese Leads insgesamt.

Ein Modell, das Ergebnisse erklären soll, ist noch kein Score. Im Modell von Virtanen et al.³ hatte die Kontrollvariable Lead-Dauer einen deutlichen Effekt auf den Abschlusserfolg („The third control variable, lead-time, had a clear effect on sales closing performance.“), und die Lead-Dauer wurde in Quartalen von der Identifikation des Leads bis zu seinem Gewinn, Verlust oder Abbruch gemessen („lead time was calculated in quarters from lead identification until the deal was won, lost, or canceled“). Das passt zum Ziel der Autoren, Ergebnisse zu erklären. Die Lead-Dauer kennt man aber erst, wenn ein Lead abgeschlossen ist, und ein Score, der beim Eingang des Leads gesetzt wird, kann nur nutzen, was in diesem Moment bekannt ist. Nach der Lesart dieses Leitfadens lässt eine Eingangsgröße, die bis zum Ergebnis gemessen wird, ein Modell beim Erklären besser aussehen, als ein Score, der beim Eingang des Leads gesetzt wird, beim Vorhersagen sein könnte.

Ein Score kann sich auch selbst bestätigen. Werden Leads mit niedrigem Score weniger nachverfolgt, konvertieren weniger von ihnen, und die nächste Datenrunde sagt, die niedrigen Scores hätten gestimmt. Die Ergebnisse von Sabnis et al.² machen das plausibel, weil

die Zeit der Vertriebsmitarbeitenden mit ihrer Bewertung der Vorauswahl zusammenhing, aber keine Quelle hier misst es. Erst die Aufzeichnung, wer welchen Lead wann nachverfolgt hat, trennt einen schwachen Lead von einem vernachlässigten. Das ist eine Überlegung dieses Leitfadens und ungeprüft.

Welche Angaben gehören zu einem Score, bevor er an den Vertrieb geht?

§ 7 von 8

Sechs, in einem Eintrag je Score: das vorhergesagte Ergebnis und seine Definition, das Zeitfenster, die Grundgesamtheit und was ausgeschlossen wurde, die Eingangsgrößen und ob jede zum Zeitpunkt der Bewertung bekannt war, der Vergleich mit der Basisrate an Leads, an die der Score nicht angepasst wurde, und die Aufzeichnung der Nachverfolgung. Die Felder folgen der Zuordnung der Ergebnisse und dem Prüffenster bei Marvasti et al.¹, den Ausschlüssen und den Vergleichen mit Zufall und häufigstem Ergebnis bei Virtanen et al., der Prüfung für Rangfolgen, die dieser Leitfaden ergänzt, sowie den Befunden zur Nachverfolgung bei Sabnis et al.²

Abbildung 2 Ein Blatt für einen Lead-Score

Ergebnis, Zeitfenster, Grundgesamtheit, Eingangsgrößen, Prüfung und Nachverfolgung festhalten, bevor ein Score entscheidet, wer angerufen wird.

Ergebnis	Zeitfenster	Grundgesamtheit	Eingangsgrößen	Prüfung	Nachverfolgung
Welches Ereignis, wie definiert?	Bis wann nach der Bewertung?	Welche Leads, welche ausgeschlossen?	Bei der Bewertung bekannt?	An neuen Leads besser als die Basisrate?	Wer hat wann gehandelt?

Ein Score, der an Leads, die er nicht kennt, nicht besser abschneidet als die Basisrate, sollte nicht entscheiden, wer angerufen wird.

Eine Zeile je Score. Den Score an Leads prüfen, die später eingingen als die, mit denen er festgelegt wurde. Ohne die Spalte zur Nachverfolgung sehen ein niedriger Score und ein vernachlässigter Lead gleich aus.

Quelle Entscheidungsblatt des Autors. Die Felder für Ergebnis und Zeitfenster folgen Marvasti et al. (2021), S. 131; das Feld für die Grundgesamtheit folgt Marvasti et al. (2021), S. 131, und Virtanen et al. (2015); die Felder für Eingangsgrößen und Prüfung folgen Virtanen et al. (2015) und Marvasti et al. (2021), S. 130 bis 132, die Prüfung für Rangfolgen ergänzt dieser Leitfaden; das Feld für die Nachverfolgung folgt Sabnis et al. (2013), S. 55, 59 und 63 bis 64. Die Felder sind Eingaben der Lesenden; es werden keine Unternehmensdaten geliefert.

Ein Beispiel ohne Unternehmensdaten: Ein B2B-Softwareanbieter bewertet die Unternehmen, die einen kostenlosen Testzugang starten, am vierzehnten Tag der Testphase. Der Eintrag nennt das Ergebnis (ein bezahlter Vertrag, kein vereinbarter Vertriebsstermin), das Zeitfenster (bis Tag 90 der Testphase, also 76 Tage nach der Bewertung), die Grundgesamtheit (Testzugänge von geschäftlichen E-Mail-Domains, ohne Bestandskunden, Partner und eigene Beschäftigte) und die Eingangsgrößen (Unternehmensgröße und Branche aus dem Anmeldeformular sowie die Produktnutzung in den ersten 14 Tagen, nichts, was später erfasst wird). Er prüft den Score an Testzugängen, die nach den Monaten begannen, mit denen die Gewichte festgelegt wurden, und vergleicht den Anteil bezahlter Verträge im obersten Zehntel der Scores mit dem Anteil unter allen Testzugängen. Er hält fest, welche Testzugänge der Vertrieb wann kontaktiert hat, damit ein

niedriger Score nicht durch Vernachlässigung bestätigt wird. Um zu erfahren, ob es die Conversion verändert, die am höchsten bewerteten Testzugänge zuerst anzurufen, und nicht nur, ob der Score sie vorhersagt, würde der Anbieter unter den Testzugängen über dem Schwellenwert losen, die eine Hälfte sofort an den Vertrieb geben und die andere nach dem üblichen Ablauf, und am Ende der vollen 90 Tage den Anteil mit bezahltem Vertrag in beiden Hälften vergleichen. Das sind Beispiele des Autors und ungeprüft.

Was müsste ein Softwareteam noch prüfen?

§ 8 von 8

Wie groß all das in einem Softwaregeschäft ist. Sabnis et al.² befragten Vertriebsmitarbeitende in vier Unternehmen, die seit Jahrzehnten zu den großen Anbietern ihrer Branche gehören („major players in their industries“), zu ihren Zeitanteilen und warnen, dass ihre Schlüsse für kleinere oder jüngere Unternehmen, deren Vertriebsprozesse sich noch entwickeln, nicht gelten könnten („our conclusions might not apply to smaller or younger firms whose sales processes are still evolving“). Marvasti et al.¹ testeten ihr Modell an simulierten Unternehmen und an den spärlichen Surfdaten eines Anbieters, die nur zusammengefasst vorlagen, sodass sich das Surfverhalten nicht nach Produkten oder Entscheidern trennen ließ („the browsing data is available only at an aggregated level, where it is not possible to separate browsing associated with different products or decision-makers“). Virtanen et al.³ untersuchten die CRM-Daten eines IT-Lösungsanbieters und konnten aus Gründen der Vertraulichkeit nicht angeben, wie viele Leads gewonnen, verloren oder abgebrochen wurden („the total amount of leads won, lost, and canceled cannot be mentioned in this paper“).

Nach der Lesart dieses Leitfadens lassen sich zwei Dinge mit einiger Sicherheit auf ein Softwareteam übertragen, weil sie aus dem folgen, was ein Score ist, und nicht aus einer einzelnen Stichprobe: Ein Score muss Ergebnis, Zeitfenster und Grundgesamtheit benennen, und was er leistet, sagt wenig, bevor es an Leads, an die er nicht angepasst wurde, an der Basisrate gemessen wird. Ein Drittes ist eine Hypothese, die zu prüfen bleibt, weil die Autoren warnen, dass ihre Schlüsse für kleinere oder jüngere Unternehmen nicht gelten könnten: dass die Zeit, die Menschen den erhaltenen Leads widmen, damit zusammenhängt, wie gut sie die Vorauswahl bewerten, durch die diese Leads kamen. Offen lassen die Quellen, ob Regeln oder Modelle Leads besser ordnen, wie stark das Handeln nach einem Score die Conversion verändert und ob Vertriebsmitarbeitende dem Score eines Modells mehr oder weniger trauen als einer Regel, die sie nachlesen können.

 Ende des Beitrags

Referenzen

Zitate stehen in der Sprache des Originals; die deutschen Wiedergaben im Fließtext sind sinngemäß und stammen vom Autor.

- 1 Marvasti, N. B., Huhtala, J.-P., Yousefi, Z. R., Vaniala, I., Upreti, B., Malo, P., Kaski, S., & Tikkanen, H. (2021). Is this company a lead customer? Estimating stages of B2B buying journey. *Industrial Marketing Management*, 97, 126–133. <https://doi.org/10.1016/j.indmarman.2021.06.003>
- 2 Sabnis, G., Chatterjee, S. C., Grewal, R., & Lilien, G. L. (2013). The sales lead black hole: On sales reps' follow-up of marketing leads. *Journal of Marketing*, 77(1), 52–67. <https://doi.org/10.1509/jm.10.0047>
- 3 Virtanen, T., Parvinen, P., & Rollins, M. (2015). Complexity of sales situation and sales lead performance: An empirical study in business-to-business company. *Industrial Marketing Management*, 45, 49–58. <https://doi.org/10.1016/j.indmarman.2015.02.024>

Zitierweise

Isoglu, S. (2026, 18. Oktober). Was ist Lead-Scoring? Sinan Isoglu.
<https://isoglu.com/de/insights/journal/was-ist-lead-scoring/>
Online-Fassung: <https://isoglu.com/de/insights/journal/was-ist-lead-scoring/>

Sinan Isoglu, MBA (Quantic)

Führungskraft für Umsatzwachstum, Dozent und Doktorand
© 2026 Sinan Isoglu · isoglu.com · Alle Rechte vorbehalten