

MANAGEMENT REVIEW

Management ohne Theater.

KI

BEDARFSPROGNOSE: WAS MASCHINELLES LERNEN ÄNDERTE

Ein Prognosewettbewerb mit Walmart-Absätzen, die meisten Teams hinter einfachen Vergleichsmethoden, was zweimal gewann, wo die Gewinne enden, wie man Prognose und Handkorrektur misst, und eine Karte.



22,4 %

genauer als die beste statistische Vergleichsmethode: der Sieger des M5-Wettbewerbs mit Walmart-Absätzen, 2020.

MAKRIDAKIS ET AL., 2022

IN DIESER AUSGABE

TITELTHEMA

**Zehn Filialen, 42.840
Reihen**

SEITE 03

WAS DIE FORSCHUNG SAGT

Wo die Gewinne enden

SEITE 06

WERKZEUG DER AUSGABE

**Die Karte zur
Prognoseprüfung**

SEITE 08

NR. 77 · KI

In dieser Ausgabe

Eine Bedarfsprognose entscheidet, was bestellt wird, wie viele Leute im Dienstplan stehen und wie viel Bestand im Regal wartet. Maschinelles Lernen verspricht bessere Prognosen. Diese Ausgabe fragt, was die größten offenen Prognosewettbewerbe dazu fanden, was es änderte und was nicht.

Im M5-Wettbewerb 2020 prognostizierte der Sieger Walmarts Absätze um 22,4 % genauer als die exponentielle Glättung, doch nur 7,5 % der Teams schlugen diese Vergleichsmethode. In M4 wie in M5 gewannen Kombinationen; maschinelles Lernen gewann, wenn es über viele verwandte Reihen zugleich lernte. Für den Gesamtabsatz waren die Gewinne groß, für ein Produkt in einer Filiale klein. Einfache Vergleichsmethoden zeigen, ob eine Methode half, und vier Unternehmen zeigen, wann die Handkorrektur eines Planers half.

Stiven Janaqi HERAUSGEBER · STIVEN CATALYST

03 TITELTHEMA
Zehn Filialen, 42.840 Reihen

04 DIE ZAHLEN
Die meisten schlugen die einfachen nicht

05 DAS MODELL
Was gewann, zweimal

06 WAS DIE FORSCHUNG SAGT
Wo die Gewinne enden

07 WIE MAN ES MISST
Messen gegen die naive Prognose

08 WERKZEUG DER AUSGABE
Die Karte zur Prognoseprüfung

09 QUELLEN
Quellen und Methode

SO LESEN SIE DIESE AUSGABE

Zahl

Jede Zahl hat Quelle und Jahr unten auf ihrer Seite.

Unsere Lesart

Wo die Redaktion deutet und nicht die Forschung, steht es dabei.

Praxis

Schritte und Karte sind Vorschläge zum Ausprobieren, keine Forschungsergebnisse.

TITELTHEMA

Zehn Filialen, 42.840 Reihen

2020 gab der M5-Wettbewerb Prognostikern mehr als fünf Jahre tägliche Stückabsätze von Walmart: 3.049 Produkte in zehn Filialen in Kalifornien, Texas und Wisconsin. Aufgabe war, die nächsten 28 Tage vorherzusagen.

42.840

Reihen auf 12 Ebenen: jedes Produkt je Filiale, summiert nach Abteilung, Kategorie, Filiale und Bundesstaat

30.490

Prognosen auf der untersten Ebene, ein Produkt je Filiale; der Rest daraus summiert

Das Siegerteam war 22,4 % genauer als die beste statistische Vergleichsmethode, die exponentielle Glättung. Nur die ersten fünf kamen über 20 %, alle 50 besten über 14 %. Der Sieger mittelte gleich gewichtet viele LightGBM-Modelle, ein Verfahren aus Entscheidungsbäumen, jedes über alle Reihen einer Filiale, Filialkategorie oder Filialabteilung trainiert.

UNSERE LESART

Ein Teil des Gewinns kam von der Saisonmodellierung und äußeren Informationen wie Preisen; der Rest, meinen die Veranstalter, vom Lernen über Tausende verwandter Reihen zugleich.

Quellen: Spyros Makridakis et al., International Journal of Forecasting 38(4), The M5 competition: Background, organization, and implementation, 2022 (über Abstract; the intermittency shares from a later paper); Spyros Makridakis et al., International Journal of Forecasting 38(4), M5 accuracy competition: Results, findings, and conclusions, 2022 (über The authors' preprint, 2020)

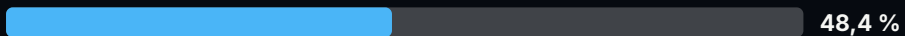
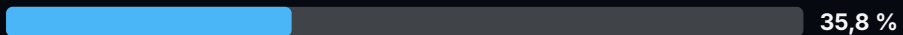
Genauigkeit als WRMSSE: Fehler jeder Reihe skaliert am Fehler einer naiven Prognose und gewichtet nach ihrem Umsatz in Dollar. Die Ergebnisarbeit lag als Vorabfassung der Autoren (2020) vor, die Arbeit zu den Daten nur als Abstract.

DIE ZAHLEN

Die meisten schlugen die einfachen nicht

Neben den Einsendungen ließen die Veranstalter Vergleichsmethoden laufen, von der naiven Prognose, die den letzten Wert wiederholt, bis zur exponentiellen Glättung. Von den Teams der Schlusswertung schlug jede davon dieser Anteil:

TEAMS, DIE JEDE VERGLEICHSMETHODE SCHLUGEN, M5, 2020

Naive Prognose**Saisonal naive Prognose****Exponentielle Glättung, die beste Methode**

Hätte jedes Team seinen besten Versuch eingereicht, hätten 63,7 %, 48,8 % und 12,2 % sie geschlagen. Die Veranstalter vermuten irreführende Validierungswerte: Viele Teams wählten nicht die beste Methode, die sie gebaut hatten.

UNSERE LESART

Ein komplexes Modell ist leicht gebaut und schwer geprüft. Ohne einfache Vergleichsmethode daneben weiß niemand, ob es half.

Quelle: Spyros Makridakis et al., International Journal of Forecasting 38(4), 2022 (über The authors' preprint, 2020)

Endgültige Einsendungen. Die Veranstalter fügen an, dass viele Teams hinter der exponentiellen Glättung das Interesse verloren haben könnten; die 7,5 % sind daher womöglich etwas verzerrt.

DAS MODELL

Was gewann, zweimal

Zwei Jahre vor M5 verlangte der M4-Wettbewerb Prognosen für 100.000 Zeitreihen. Seine Lehren und die von M5 weisen in dieselbe Richtung, mit einer Änderung.

M4, 2018

12 der 17 genauesten Methoden waren Kombinationen

der Sieger, ein Hybrid aus Statistik und neuronalem Netz, rund 10 % besser als eine Vergleichskombination

sechs reine Methoden des maschinellen Lernens: keine schlug die Kombination, eine die naive

M5, 2020

LightGBM in den meisten Spitzenbeiträgen

Kombinationen vieler Modelle

ein Modell, das über viele Reihen lernt

Preise, Aktionen und Ereignisse als Eingaben

UNSERE LESART

Maschinelles Lernen gewann, wo es viele verwandte Reihen und Informationen über die Absatzhistorie hinaus gab, wie in M5. Reines maschinelles Lernen in M4 hatte nicht gewonnen.

Quellen: Spyros Makridakis et al., International Journal of Forecasting 34(4), 2018 (über Abstract in library records); Spyros Makridakis et al., International Journal of Forecasting 38(4), 2022 (über The authors' preprint, 2020)

Die 10 % in M4 beziehen sich auf sMAPE, einen prozentualen Fehler. In M5 war exponentielle Glättung mit Aktionen und Sondertagen 6 % genauer als ohne, ARIMA mit solchen Eingaben 13 %.

WAS DIE FORSCHUNG SAGT

Wo die Gewinne enden

Die 42.840 Reihen liegen auf 12 Ebenen, vom gesamten Walmart-Absatz in den Daten bis zu einem Produkt in einer Filiale. Die Veranstalter verglichen die besten Methoden auf jeder Ebene mit der besten Vergleichsmethode.

DURCHSCHNITTLICHER GEWINN DER BESTEN METHODEN GEGENÜBER DER BESTEN VERGLEICHSMETHODE, NACH EBENE (2020)



Für ein Produkt in einer Filiale waren 73 % der 30.490 Reihen sporadisch: Tage ohne Verkauf, dann ein paar. Dort, so die Veranstalter, hatten die üblichen Methoden für sporadischen Bedarf noch einen gewissen Wert.

UNSERE LESART

Die Prognose, die das Regal füllt, bleibt die unsichere. Bestand, Personal und Kapazität für den Fehler planen, nicht nur für die Zahl.

Quellen: Spyros Makridakis et al., International Journal of Forecasting 38(4), M5 accuracy competition: Results, findings, and conclusions, 2022 (über The authors' preprint, 2020); Spyros Makridakis et al., International Journal of Forecasting 38(4), The M5 competition: Background, organization, and implementation, 2022 (über Abstract; the intermittency shares from a later paper)

Gerundete Durchschnitte, gelesen in der Vorabfassung der Ergebnisarbeit. Auf den drei untersten Ebenen lag der Gesamtsieger unter den besten 50 auf Platz 11 bis 13.

[Mehr im Essay](#)

Tage mit hohem Volumen: der Standard unter Druck

WIE MAN ES MISST

Messen gegen die naive Prognose

Rob Hyndman und George Athanasopoulos nutzen im ganzen Lehrbuch vier einfache Methoden als Vergleich; die einfachste, die naive Prognose, wiederholt den letzten Wert. Andere Methoden werden mit ihnen verglichen.

DREI FEHLERMASSE

01	MAE, mittlerer absoluter Fehler	In Stück. Ihn zu minimieren führt zum Median, RMSE zu minimieren zum Mittelwert.
02	MAPE, mittlerer absoluter prozentualer Fehler	Unendlich oder undefiniert an einem Tag ohne Absatz, extrem nahe null.
03	MASE, mittlerer absoluter skalierten Fehler	Unter 1, wenn die Prognose die naive auf vergangenen Daten schlägt.

In vier Unternehmen der Lieferkette, bei über 60.000 Prognosen, änderten Planer im Schnitt 75 % der Systemprognosen. Große Korrekturen halfen mehr, kleine schaden oft. Korrekturen nach oben halfen viel seltener und gingen öfter in die falsche Richtung.

UNSERE LESART

Jede Hand an der Prognose messen: das System gegen die naive, die Korrektur gegen das System.

Quellen: Rob J. Hyndman & George Athanasopoulos, OTexts, 2021; Robert Fildes et al., International Journal of Forecasting 25(1), 2009 (über Abstract and summaries); Robert Fildes & Paul Goodwin, Foresight 8, 2007

Maße wie bei Hyndman & Athanasopoulos (2021). Fildes et al.: drei Hersteller mit Monats-, ein Händler mit Wochenprognosen; die Autoren sehen im Hang nach oben Optimismus.

WERKZEUG DER AUSGABE

Die Karte zur Prognoseprüfung

Eine Karte pro Produktfamilie und Monat: die naive Prognose neben der des Systems, jede Handkorrektur mit Grund, und ob der Fehler zu einer Seite neigt.

01 Reihe und Ebene Produktfamilie, Filiale oder Region; wie viele Tage ohne Absatz

02 Vergleich naiv oder saisonal naiv: sein Fehler im selben Zeitraum

03 Systemprognose Fehler in Stück (MAE) und gegen den Vergleich (MASE)

04 Handkorrekturen wer was änderte, nach oben oder unten, um wie viel, und warum

05 Wert der Korrektur Fehler nach der Korrektur gegen die Systemprognose davor

06 Fehlerneigung Monate über gegen unter dem Absatz; die Fehler in einer Regelkarte

07 Nächster Monat behalten, Methode oder Daten ändern, nutzlose Korrekturen lassen

Quellen: Rob J. Hyndman & George Athanasopoulos, OTexts, 2021; Robert Fildes et al., International Journal of Forecasting 25(1), 2009 (über Abstract and summaries)

Eine Praxis, die die Redaktion vorschlägt, nach Hyndman & Athanasopoulos (2021) und Fildes et al. (2009). Zum Monatsplan selbst siehe Nr. 59.

Tool öffnen: Sigma & Regelkarte

stivencatalyst.com/de/tools/sigma-control-chart/

QUELLEN UND METHODE

Jede Zahl hat eine Quelle.

Die Zahlen dieser Ausgabe stammen aus den folgenden Quellen. Das Jahr zeigt, wie aktuell jede ist.

M5 accuracy competition: Results, findings, and conclusions

Spyros Makridakis et al., International Journal of Forecasting 38(4), 2022 · über The authors' preprint, 2020

<https://doi.org/10.1016/j.ijforecast.2021.11.013>

The M5 competition: Background, organization, and implementation

Spyros Makridakis et al., International Journal of Forecasting 38(4), 2022 · über Abstract; the intermittency shares from a later paper

<https://www.sciencedirect.com/science/article/pii/S0169207021001187>

The M4 Competition: Results, findings, conclusion and way forward

Spyros Makridakis et al., International Journal of Forecasting 34(4), 2018 · über Abstract in library records

<https://doi.org/10.1016/j.ijforecast.2018.06.001>

Forecasting: Principles and Practice, 3rd edition

Rob J. Hyndman & George Athanasopoulos, OTexts, 2021

<https://otexts.com/fpp3/>

Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning

Robert Fildes et al., International Journal of Forecasting 25(1), 2009 · über Abstract and summaries

<https://doi.org/10.1016/j.ijforecast.2008.11.010>

Good and Bad Judgment in Forecasting: Lessons from Four Companies

Robert Fildes & Paul Goodwin, Foresight 8, 2007

https://forecasters.org/wp-content/uploads/Good-and-Bad-Judgment-in-Forecasting_Issue8.pdf

REDAKTIONELLE METHODE

Jede Zahl wurde auf Jahr, Herausgeber und genaue Bedeutung geprüft. Wo die Seite des Herausgebers nicht erreichbar war, wurde sie an unabhängigen Zusammenfassungen geprüft und mit „über“ markiert. Die Deutung der Redaktion ist als „Unsere Lesart“ markiert. Zahlen, die sich nicht bestätigen ließen, stehen nicht in der Ausgabe.



MANAGEMENT REVIEW

Management ohne Theater.

Jede Ausgabe eine Führungsfrage, geprüft an der besten Forschung.

ALLE AUSGABEN

stivencatalyst.com/de/magazine/management-review.html