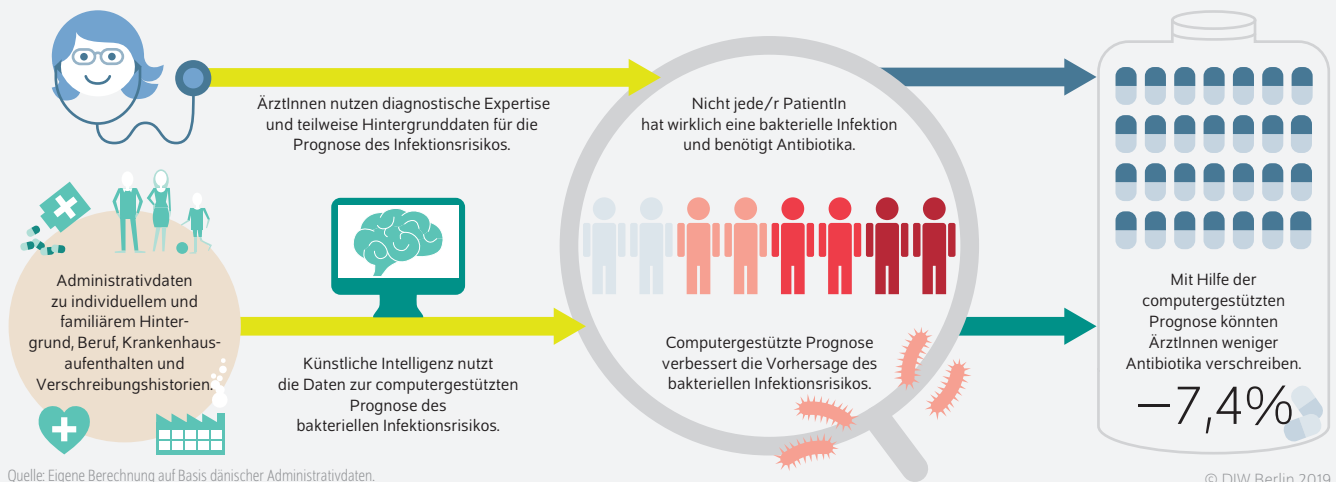


Künstliche Intelligenz und Daten können bei der Eindämmung von Antibiotikaresistenzen helfen

Von Michael A. Ribers und Hannes Ullrich

- Untersucht wird, ob Daten und maschinelles Lernen die ärztliche Verschreibungspraxis verbessern können
- Unter Nutzung umfassender administrativer und medizinischer Daten aus Dänemark werden Methoden des maschinellen Lernens und ein ökonomisches Modell verwendet
- Dies ermöglicht die Entwicklung von Verschreibungsregeln zur Reduktion von Fehlverschreibungen von Antibiotika
- Die Gesamtzahl an Verschreibungen für Harnwegsinfekte kann durch die vorgeschlagenen Politikmaßnahmen um 7,4 Prozent verringert werden, ohne die Anzahl behandelter bakterieller Infektionen zu verringern
- In Deutschland kann das große Potential dieser Methode erst durch weitere Bemühungen bei der Digitalisierung von Informationen verwirklicht werden

Maschinelles Lernen kann bakterielle Infektionen vorhersagen und könnte ärztliche Verschreibungen von Antibiotika bei Harnwegsinfektionen um 7,4 Prozent senken



ZITAT

„Unsere Ergebnisse zeigen: Der Einsatz künstlicher Intelligenz kann ein wichtiger Baustein sein im Kampf gegen die Verbreitung von Antibiotikaresistenzen. Denn nur wenn Antibiotika zweckdienlich genutzt, also verschrieben werden, kann ihr therapeutischer Nutzen erhalten bleiben.“

— Hannes Ullrich, Autor der Studie —

MEDIATHEK



Audio-Interview mit Hannes Ullrich
www.diw.de/mediathek

Künstliche Intelligenz und Daten können bei der Eindämmung von Antibiotikaresistenzen helfen

Von Michael A. Ribers und Hannes Ullrich

ABSTRACT

Eine zentrale Strategie, um der Zunahme von Antibiotikaresistenzen entgegenzutreten, ist die Verbesserung der ärztlichen Verschreibungspraxis. Damit sollen Fehlverschreibungen von Antibiotika als eine Hauptursache von Antibiotikaresistenzen vermieden werden. Die zunehmende Verfügbarkeit medizinischer Daten und Methoden des maschinellen Lernens bieten die Chance, schnellere Diagnosen bereitzustellen. In diesem Beitrag wird am Beispiel von Antibiotikaverschreibungen für Harnwegsinfekte in Dänemark dargelegt, wie datenbasierte Vorhersagen die klinische Praxis vor dem Hintergrund von Antibiotikaresistenzen verbessern können. Zu diesem Zweck werden umfassende administrative und medizinische Daten zusammen mit Methoden des maschinellen Lernens und einem ökonomischen Modell verwendet, um Verschreibungsregeln für Antibiotika zu entwickeln. Dabei zeigt sich, dass die Gesamtzahl an Verschreibungen durch die vorgeschlagenen Politikmaßnahmen um 7,42 Prozent verringert werden kann, ohne die Anzahl behandelter bakterieller Infektionen zu verringern. Dies demonstriert das große Potential dieser Methode. In Deutschland kann dieses Potential jedoch erst durch weitere Bemühungen bei der Digitalisierung von Informationen verwirklicht werden. Eine wichtige Rolle spielt die Verknüpfung oft dezentral erhobener und gespeicherter Informationen, die für die IT-Systeme von Arztpraxen und Krankenhäusern verfügbar sein müssen.

In den 1940er Jahren revolutionierte die Entwicklung von Penicillin zur therapeutischen Nutzung die Medizin, ermöglichte neue medizinische Eingriffe und rettete seither Millionen von Menschenleben. Mit der immer stärkeren Verwendung von Antibiotika für medizinische und nicht-medizinische Zwecke – wie zum Beispiel in der Agrarindustrie – trat in den letzten Jahrzehnten das Problem wachsender Antibiotikaresistenzen stärker in den Vordergrund. Multiresistente Bakterien – darunter die weit verbreiteten *Escherichia coli*, *Klebsiella pneumoniae* und *Staphylococcus aureus* – werden wieder zu lebensgefährlichen Risiken. Jedes Jahr verlieren rund 700 000 Menschen weltweit ihr Leben durch nicht behandelbare Infektionen. Die Weltgesundheitsorganisation sieht Antibiotikaresistenzen als eine der größten gesundheitspolitischen Herausforderungen unserer Zeit. So werden die mit resistenten Erregern verbundenen Kosten für das Jahr 2050 auf 100 Billionen US-Dollar prognostiziert.¹ Fehlende finanzielle Anreize zur Entwicklung neuer Antibiotika stellen ein Hauptproblem für das Gesundheitswesen dar, welchem durch angebotsseitige Politikmaßnahmen – wie Subventionen oder Ausnahmeregelungen im Patentsystem – begegnet werden kann.² In der Vergangenheit haben sich jedoch gegen jedes neu entwickelte Antibiotikum Resistenzen entwickelt und dadurch ihren therapeutischen Wert über die Zeit gesenkt. Ein Hauptaugenmerk der Gesundheitspolitik liegt daher auf der ärztlichen Verschreibungspraxis, welche eine Hauptrolle für die Erhaltung des therapeutischen Nutzens vorhandener Medikamente spielt.³

Die hier vorgestellte, vom Europäischen Forschungsrat geförderte Studie am DIW Berlin zeigt am Beispiel von Dänemark, wie eine Politikmaßnahme unter Verwendung umfassender Individualdaten und maschinellen Lernens bei vermuteten Harnwegsinfekten die Anzahl unnötiger Verschreibungen senken und die Anzahl zielführender

¹ Jim O'Neill (2016): Tackling drug-resistant infections globally: final report and recommendations. Review on Antimicrobial Resistance. HM Government and Wellcome Trust: UK (online verfügbar).

² Vgl. Ramanan Laxminarayan et al. (2013): Antibiotic resistance – the need for global solutions, *The Lancet Infectious Diseases*, 13(12), 1057-1098.

³ Vgl. Eili Y. Klein et al. (2018): Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. *Proceedings of the National Academy of Sciences*, 115(15), E3463-E3470.

Verschreibungen erhöhen kann.⁴ Insbesondere wird gezeigt, dass datenbasierte Vorhersagen so sicher zwischen bakteriellen und nicht-bakteriellen Ursachen unterscheiden können, dass eine Reduktion von rund 7,42 Prozent der getätigten Verschreibungen möglich ist, ohne die Anzahl behandelter bakterieller Infektionen zu verringern. Da dieses Ergebnis auf einer Umverteilung von Verschreibungen basiert, werden auch damit verbundene Risiken diskutiert. Abschließend werden Überlegungen dargestellt, welche Herausforderungen sich für vergleichbare Maßnahmen in Deutschland ergeben.

Verschreibungspraxis begünstigt Entwicklung von Resistenzen

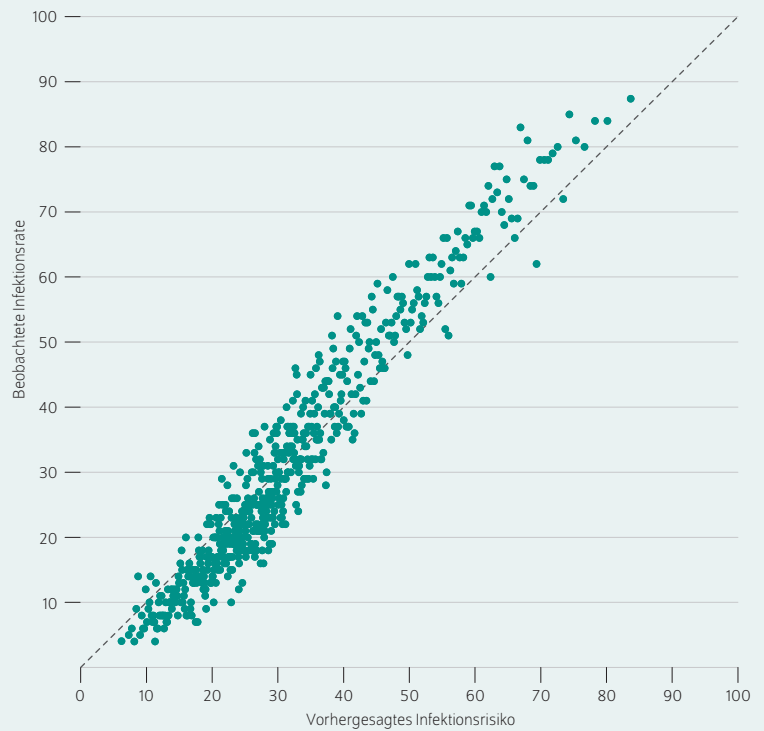
Neben mangelnder Hygiene und dem Einsatz von Antibiotika in der Landwirtschaft ist die Antibiotikaeinnahme von Menschen eine Hauptursache für die Entwicklung und Verbreitung von Bakterien, die gegen bestimmte oder mehrere Wirkstoffe resistent sind. Je mehr Bakterien antibiotischen Wirkstoffen ausgesetzt sind, desto mehr entwickeln sich resistente Bakterien, solange sie die Konfrontationen mit diesen Substanzen überleben. Resistenzfördernd sind sowohl unnötige Überverschreibungen, die Bakterien Wirkstoffen aussetzen, als auch Unterverschreibungen, die ein Überleben und eine Weiterentwicklung möglicherweise resistenter Bakterien erlauben.

In einer typischen allgemeinärztlichen Behandlungssituation, in welcher eine Patientin oder ein Patient erstmalig mit Symptomen einer Infektion in der Praxis erscheint, müssen Ärztinnen und Ärzte zwei Probleme lösen. Zuerst muss eine Diagnose erstellt werden, die eine Ursachenabschätzung beinhaltet. Im zweiten Schritt muss der erwartete positive Effekt im Falle einer Behandlung gegen die erwartete Entwicklung von Resistenzen abgewogen werden. Durch aktuell verfügbare Diagnosemethoden in Allgemeinarztpraxen ist die Diagnoseunsicherheit insbesondere bei Erstuntersuchungen hoch.⁵ Es können zwar detaillierte, mikrobiologische Labortests, die potentielle Erreger identifizieren und Informationen über ihre Resistenzprofile erstellen, angefordert werden. Diese Informationen sind in der Regel aber erst nach einigen Tagen verfügbar, was oft der Dauer einer Behandlungseinheit mit den Antibiotika entspricht. Patientinnen und Patienten wünschen bzw. benötigen jedoch direkt eine Behandlung. Diese Entscheidung unter Unsicherheit führt zu Über- und Unterverschreibungen.

Die Entwicklung von Schnelldiagnoseverfahren wird als ein wichtiges Mittel zur Eindämmung unnötiger Verschreibungen beschrieben.⁶ Eine potentielle Rolle wird hier bereits datenbasierten Verfahren eingeräumt.⁷ Sofortige, detaillierte

Abbildung 1

Im Labor beobachtete bakterielle Infektionsrate nach vorhergesagtem bakteriellen Risiko in Prozent
 Jeder Punkt steht für 100 Beobachtungen



Quelle: Eigene Berechnungen auf Basis dänischer Administrativdaten.

© DIW Berlin 2019

Die Nähe der Punkte zur Diagonalen ist ein Zeichen für die hohe Vorhersagequalität.

Diagnosen, die über spezifische Bakterien und die Wirksamkeiten von einzelnen Antibiotika Aufschluss geben, werden in absehbarer Zeit eine Herausforderung bleiben. Allerdings ist es für die große Masse an Verschreibungsentscheidungen bereits hilfreich, sofortige Informationen über die Wahrscheinlichkeit einer bakteriellen Infektion bereitzustellen. Auf dieser Grundlage kann fundierter entschieden werden, ob eine Antibiotikabehandlung bis zum Eintreffen detaillierter Testergebnisse aufgeschoben werden kann oder direkt begonnen werden sollte.

Daten und maschinelles Lernen können die Verschreibungspraxis verbessern

Ärztinnen und Ärzte haben grundsätzlich vielfältige Hintergrundinformationen über Patientinnen und Patienten, die ihnen bei der ersten Diagnose helfen können. So wissen sie, ob sie eine Person vor sich haben, die eher bakteriellen Erregern ausgesetzt ist (zum Beispiel, weil sie in der Schweine-schlachtung und nicht im Hochbau arbeitet). Genauso könnten sie individuelle Behandlungs- und Diagnosehistorien erheben und vieles mehr. Diese Informationen sind von Natur aus multidimensional und umfassend, so dass ihre Nutzung mit erheblichem Aufwand verbunden ist. Ihre

4 Vgl. Michael A. Ribers und Hannes Ullrich (2019): Battling Antibiotic Resistance: Can Machine Learning Improve Prescribing? DIW Diskussionspapier Nr. 1803.

5 Vgl. Angela M. Caliendo et al. (2013): Better tests, better care: improved diagnostics for infectious diseases. Clinical Infectious Diseases, 57(suppl_3), 139-170.

6 O'Neill (2016), a. a. O.

7 O'Neill (2016), a. a. O., Seite 37.

Kasten 1

Maschinelles Lernen für die Politik

Maschinelles Lernen trifft Vorhersagen rein basierend auf Korrelationen des vorherzusagenden Ergebnisses mit beobachtbaren Informationen. Wird eine Variable gefunden, die wichtig für die Vorhersage eines Ergebnisses ist, im vorliegenden Kontext zum Beispiel der Zeitpunkt und die Art der letzten Antibiotikaveranschreibung, trägt diese substantiell zur Vorhersagequalität bei. Dies erlaubt jedoch nicht die Schlussfolgerung einer Kausalität. Die Ursache für eine aktuelle bakterielle Infektion muss nicht in einer vorherigen Behandlungsentscheidung liegen. Vielmehr ist wahrscheinlich, dass ein in den Daten nicht beobachtbarer Faktor, wie zum Beispiel ein erhöhter Konsum kontaminierter Lebensmittel, sowohl zur letzten Behandlungsentscheidung als auch zur aktuellen Infektion führte. Für die Vorhersagequalität ist diese Unterscheidung irrelevant, da vergangene Behandlungsentscheidungen diese Informationen enthalten.

ÖkonomInnen untersuchen traditionell kausale Zusammenhänge, um zum Beispiel die Auswirkungen einer Politikmaßnahme auf Marktergebnisse zu quantifizieren. Durch die Verfügbarkeit detaillierter, hochfrequenter Daten, die Entwicklung von Algorithmen und die zunehmenden Rechenkapazitäten bietet sich ein verstärkter Fokus auf Politikmaßnahmen an, die zunächst ein Vorhersageproblem beinhalten.¹ Dies kann wie in der vorliegenden Studie

¹ Vgl. Jon Kleinberg et al. (2015): Prediction policy problems, *American Economic Review Papers & Proceedings*, 105(5), 491-95. Helen Margetts und Cosmina Dorobantu (2019): Rethink government with AI. *Nature* 568, 163-165.

eine Abschätzung des Risikos einer bakteriellen Infektionsursache sein. In den USA hat die Schätzung der Rückfallrate von auf Kaution frei gelassenen Angeklagten viel Aufmerksamkeit erlangt, als die Risikoschätzung anhand von Daten und maschinellem Lernen diskutiert wurde.² Unter anderem konnte das Potential für eine Senkung der Kriminalitätsrate sowie für ein geringeres Maß an Diskriminierung benachteiligter Gruppen gezeigt werden.³ Die kontext-spezifische, sorgfältige Auswertung neuer datenbasierter politischer Handlungsspielräume erscheint somit als ein vielversprechendes Feld für weitere Forschung und gesellschaftliche Diskussionen. Neben der Quantifizierung des Potentials möglicher Anwendungen sind für eine Umsetzung auch ethische Dimensionen essentiell, zum Beispiel in Bezug auf Diskriminierung und Fairness, für die stets der Vergleich zum beobachtbaren menschlichen Handeln gezogen werden sollte.⁴

² Vgl. Julia Angwin et al. (2016): Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*, 23. Mai 2016 (online verfügbar, abgerufen am 9. April 2019). Sam Corbett-Davies et al. (2016): A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear. *The Washington Post*, 17. Oktober 2016 (online verfügbar, abgerufen am 9. April 2019).

³ Jon Kleinberg et al. (2018): Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1), 237-293.

⁴ Vgl. Jon Kleinberg, Sendhil Mullainathan und Manish Raghavan (2016): Inherent trade-offs in the fair determination of risk scores. *arXiv:1609.05807*.

Auswertung erfordert ein hohes Maß an Informationsverarbeitungskapazitäten, Expertise und Zeitaufwand. Methoden des maschinellen Lernens liefern ein Werkzeug, welches diese Komplexität auf ein Minimum reduzieren kann.

Die Stärke des maschinellen Lernens, als bedeutendstes Werkzeug der künstlichen Intelligenz, liegt in der Lösung von häufig wiederkehrenden und klar abgrenzbaren Problemen, über welche ein hoher Informationsgehalt in Form von digitalisierten Daten verfügbar ist (Kasten 1). Entscheidende technologische Fortschritte haben die technologische und wirtschaftliche Machbarkeit geschaffen, um auf Grundlage einer Fülle an Informationen gehaltvolle und einfach interpretierbare Vorhersagen zu treffen.⁸ Das Paradebeispiel in der Medizin ist die Bilderkennung in der Radiologie und der Augenheilkunde, wo hohe Fallzahlen, die Digitalisierung des Bildmaterials und von Auswertungen sowie die relativ einfache Problemstellung zur Alltagstauglichkeit geführt haben.⁹

In der hier vorgestellten Studie wird ein Ansatz verwendet, der einen Vergleich der Vorhersagen basierend auf Daten

und maschinellem Lernen mit der ärztlichen Verschreibungspraxis ermöglicht. Er bildet die Grundlage für datenbasierte Verschreibungsregeln, die die Verschreibungspraxis verbessern können. Konkret werden die vorhergesagten bakteriellen Risiken bei Diskrepanz zur ärztlichen Verschreibungsentscheidung für individuelle PatientInnen verwendet, um sofortige Verschreibungen für PatientInnen mit niedrigem vorhergesagtem Risiko zu verhindern und für PatientInnen mit hohem vorhergesagtem Risiko zu erwirken. Dafür werden Labortestergebnisse von Patientinnen und Patienten mit vermuteten Harnwegsinfektionen aus dem größten medizinischen Labor in Dänemark anonymisiert und mit umfassenden Administrativdaten verknüpft. Zu diesen Daten gehören die vollständigen Verschreibungs- und Behandlungshistorien, vergangene Krankenhausaufenthalte, Laborergebnisse und persönliche Eigenschaften wie Alter, Geschlecht, Berufstätigkeiten, Kommune der Wohnadresse, Haushaltsgröße und -art, Zivilstand und mehr. Damit steht für die Analyse ein Datensatz über einen Zeitraum von drei Jahren mit 95 594 Beobachtungen und 1266 Variablen zur Verfügung. Die zugrundeliegenden Daten werden in Dänemark seit geraumer Zeit flächendeckend gesammelt, zentral gespeichert und anonymisiert für Forschungszwecke zur Verfügung gestellt. Grundvoraussetzung für die hier diskutierte Anwendung ist die personen- und zeitspezifische Verknüpfbarkeit dieser reichhaltigen Informationen.

⁸ Ajay Agrawal, Joshua Gans und Avi Goldfarb (2018): Prediction machines: the simple economics of artificial intelligence. Harvard Business Press.

⁹ Vgl. Ziad Obermeyer und Ezekiel J. Emanuel (2016), Predicting the future – big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, 375(13), 1216.

Für die Vorhersagen bakterieller Ursachen wird ein sogenannter *random forest* basierend auf *Regressionsbäumen* verwendet (Kasten 2).¹⁰ Die Qualität der erzielten Vorhersagen wird anhand der durchschnittlichen Rate beobachteter bakterieller Infektionen für unterschiedliche Werte des vorhergesagten Risikos einer bakteriellen Infektion gezeigt (Abbildung 1). Jeder abgebildete Punkt steht für 100 individuelle Beobachtungen, für welche die durchschnittliche Rate berechnet wird. Im Idealfall sind alle Punkte auf der Diagonalen. Die Nähe der Punkte zur Diagonalen ist ein Zeichen für die hohe Vorhersagequalität, die der *random forest* anhand der verwendeten Administrativdaten erzielt.

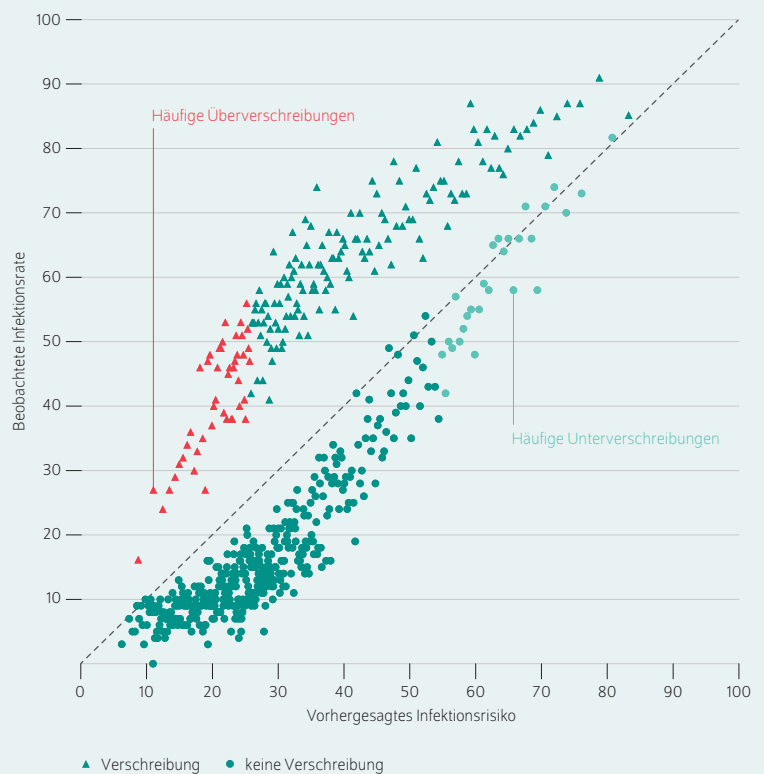
Entscheidend ist aber nicht die reine Vorhersagequalität, sondern ihr Vergleich mit ärztlichen Verschreibungsentscheidungen. Haben Ärztinnen und Ärzte bessere Informationen, als in den vom Algorithmus verwendeten Daten enthalten sind, oder verfügen sie über kontextspezifische Expertise, wäre es nicht überraschend, wenn die vorhergesagten Risiken keinen zusätzlichen Informationsgehalt hätten. So zeigt sich tatsächlich, dass nicht beobachtbare Informationen und Expertise dazu führen, dass PatientInnen mit einer Antibiotikaverschreibung (*behandelt*) – gegeben das vorhergesagte Risiko – eine höhere Rate bakterieller Infektionen haben als PatientInnen ohne Verschreibungen (*nicht-behandelt*) (Abbildung 2). Hierfür sind Diagnostikmethoden maßgeblich, die Ärzten und Ärztinnen Hinweise über das Vorliegen einer bakteriellen Infektion noch am selben Tag ermöglichen, welche jedoch ohne digitalisierte Erfassung dem Algorithmus nicht zur Verfügung gestellt werden können. Eine nähere Betrachtung in der Studie zeigt deutliche Unterschiede in der Fähigkeit der Ärztinnen und Ärzte, gemäß dem vorhergesagten Risiko Verschreibungen bei Fällen mit höherem Risiko zu verabreichen. Diese Beobachtung weist auf Unterschiede in verfügbaren Diagnosemethoden, Expertise und Verschreibungspräferenzen hin.

Es ist zu erkennen, dass eine relevante Zahl an Verschreibungen an PatientInnen mit geringem vorhergesagtem Risiko gegeben werden (Abbildung 2, rote Dreiecke). Ebenso erhält eine relevante Zahl an Personen mit hohem vorhergesagtem Risiko keine Verschreibung (Abbildung 2, hellgrüne Punkte). Der in der Studie untersuchte Vorschlag einer Maßnahme zur Reduktion von Überverschreibungen setzt hier an. Unter der Bedingung, dass die Gesamtzahl an Verschreibungen für tatsächlich bakterielle Infektionen konstant bleibt, können Verschreibungen von niedrigen Risiken an hohe Risiken umverteilt werden. Dadurch kann die Gesamtzahl an Antibiotikaverschreibungen für Harnwegsinfektionen um 7,42 Prozent reduziert werden. Alternativ kann bei gleichbleibender Anzahl an Gesamtverschreibungen eine Erhöhung der Antibiotikabehandlungen von bakteriellen Infektionen um 6,38 Prozent erreicht werden.

Da es sich um Umverteilungen handelt, ist eine berechtigte Sorge, dass Verschreibungen für PatientInnen verzögert

Abbildung 2

Bakterielle Rate nach vorhergesagtem bakteriellen Risiko in Prozent mit und ohne Antibiotikaverschreibung
 Jeder Punkt und jedes Dreieck stehen für 100 Beobachtungen



Quelle: Eigene Berechnungen auf Basis dänischer Administrativdaten.

© DIW Berlin 2019

Bei roten Dreiecken und hellgrünen Punkten kann maschinelles Lernen zu einer Verbesserung der Verschreibungspraxis führen.

werden, deren gesundheitliches Risiko durch das Vorliegen einer bakteriellen Infektion besonders hoch ist. Zum Beispiel können unbehandelte Infektionen besonders bei schwangeren Frauen, die etwa ein Drittel der Behandlungssituationen darstellen, zu Komplikationen führen. Um dies zu untersuchen, wurde dieselbe Umverteilung durchgeführt, mit der Bedingung, dass die ärztlichen Entscheidungen für ebendiese Frauen unverändert bleiben müssen. Die mögliche Reduktion an sofortigen Verschreibungen ist mit 6,81 Prozent etwas niedriger, ist aber unter Berücksichtigung statistischer Fehler von den Ergebnissen, die alle PatientInnen betreffen, nicht zu unterscheiden.

Auf der anderen Seite stellt sich die Frage, ob die zugeteilten Verschreibungen an PatientInnen mit hohem vorhergesagtem Risiko überhaupt nötig sind. Betrachtet man die ärztlichen Entscheidungen für diese PatientInnen nach Erhalt der definitiven Labordiagnose mit einigen Tagen Verzögerung, bekommen 70 Prozent der PatientInnen bei Erhalt eines positiven Testergebnisses eine Antibiotikaverschreibung. Da die Rate spontaner Heilungen von Harnwegsinfektion bei knapp 30 Prozent liegt, ist davon auszugehen,

¹⁰ Vgl. Trevor Hastie, Robert Tibshirani, und Jerome Friedman (2009): The elements of statistical learning: data mining, inference, and prediction. Springer Series in Statistics, New York.

dass die durch den Algorithmus zugeteilten Verschreibungen für die große Mehrheit angebracht sind.

Fazit: Methoden des maschinellen Lernens haben großes Potential, aber ihre Nutzung steht noch vor Herausforderungen

Die Ergebnisse zeigen das Potential für datenbasierte Vorhersagen unter Verwendung von Methoden des maschinellen Lernens. So kann die verfügbare Masse und Komplexität an Daten auf eine einfache, nützliche Information reduziert werden und so einen Beitrag zur Lösung einer der größten Herausforderungen des heutigen Gesundheitssystems leisten: die effiziente Verwendung von Antibiotika vor dem Hintergrund schnell wachsender Resistenzen. Eine versuchsweise Umsetzung in der Praxis ist aufgrund der potentiellen, wenn auch in diesem allgemeinärztlichen Kontext vermutlich überschaubaren, Risiken für einzelne Patientinnen und Patienten in Zusammenarbeit mit Ärztinnen und Ärzten nötig. Gleichzeitig bietet sich hier noch ein weiteres Potential: das Kombinieren der umfassenden Hintergrundinformationen mit fall-spezifischen Informationen, die Ärztinnen und Ärzten während der Behandlung erheben. Hierbei geht es zum Beispiel um Symptome und bestimmte verhaltensbasierte Risikofaktoren. In Dänemark bestehen hierfür die technischen Voraussetzungen, da das gesamte System der Arztpraxen digitalisiert und mit zentralen Servern verbunden ist, von welchen Ärzte und Ärztinnen Patientendaten abrufen können.

Den in der Studie im dänischen Kontext hervorgebrachten Erkenntnissen liegt die Erfüllung drei wichtiger Voraussetzungen zugrunde. Diese geben Hinweise auf die zu bewältigenden Herausforderungen für vergleichbare Anwendungen in Deutschland.

Erstens ist die Verfügbarkeit digitalisierter, standardisierter, verknüpfter und zentral verfügbarer Daten von entscheidender Bedeutung für den Erfolg maschinellen Lernens. Zur systematischen Digitalisierung und Standardisierung sind Prozesse notwendig, die Daten als Nebenprodukt von Dienstleistungen im Gesundheitssystem generieren oder gezielt Informationen mit dem Ziel spezifischer Anwendungen hinterlegen. Insbesondere die Verknüpfung und zentrale Verfügbarkeit sind Vorteile des zentralisierten dänischen Gesundheitssystems, welches die Schnittstellen und die Unterschiedlichkeit der Datenquellen minimiert und flächendeckende Datenerhebungen erlaubt. In Deutschland werden Bemühungen zur Erhebung und Verknüpfung von medizinischen und Hintergrunddaten dezentral von privaten Akteuren wie Krankenhäusern oder Versicherern unternommen.¹¹ Die Nutzung für das Design und zur Umsetzung von Politikmaßnahmen ist damit nur schwer möglich. Die Probleme bei der Einführung der digitalen Gesundheitskarte haben gezeigt, wie schwer eine zentralisierte Lösung im deutschen institutionellen Kontext ist. Im Gegensatz zu

vielen dezentralen Gesundheitsdienstleistungen könnte vor dem Hintergrund gesundheitspolitischer Herausforderungen eine unbürokratische Umsetzung eines zentralen Datenpools, zumindest aber die Gewährleistung von Konnektivität zwischen Systemen, vielseitige Potentiale zur Verbesserung der Gesundheitsversorgung schaffen.

Zweitens sind die positiven Ergebnisse im Kontext von Antibiotikaverschreibungen nicht zuletzt durch die Reduzierung des Problems auf eine einfache binäre Variable zu erreichen, die die Empfehlung für eine Ja-oder-Nein-Entscheidung anzeigt. Wie bereits beschrieben liegen die größten Potentiale bei einfachen und gängigen Vorhersageproblemen, für welche vorhersagerelevante Informationen in großem Maße vorliegen. Das Scheitern von ambitionierteren Systemen wie IBM Watson bei der Lösung komplexerer Probleme machte zuletzt deutlich, dass zum heutigen Zeitpunkt vorhandene Daten, gepaart mit effizienten Algorithmen und hohen Rechenkapazitäten, nicht automatisch zu intelligenten Vorhersagen führen.¹² Aufgrund problemspezifischer Unterschiede in der Datenerhebung, im Informationsgehalt der erhobenen Daten sowie in der Bedeutung nicht-digitalisierter, für einen Algorithmus unbeobachtbarer Informationen müssen datenbasierte Systeme kontextspezifisch implementiert und ausgewertet werden. Daraus folgt, dass die künstliche Intelligenz sich noch nicht für generelle Vorhersagen über eine große Breite an Aufgaben eignet. Ihre Stärke liegt bei klar definierbaren Problemen, für welche relevante Daten in ausreichender Häufigkeit erhoben werden. Es lohnt sich also, sich kurzfristig in erster Linie nach einfachen, automatisierbaren Problemen auf die Suche zu machen und die hier liegenden Potentiale zu nutzen.

Eine wichtige dritte Herausforderung ist der fundierte und faire Vergleich von datenbasierten Vorhersagen und Entscheidungen mit menschlichen Entscheidungen, die auf Expertenwissen basieren. Für diese Vergleiche müssen analytische Ansätze herangezogen werden, die die datengenerierenden Entscheidungsprozesse verstehen und quantifizieren können, um zum Beispiel den Einfluss verzerrter Daten oder selektierter Stichproben zu berücksichtigen.¹³ In Deutschland sind daher verstärkte Investitionen in die Ausbildung exzellenter, quantitativer und multidisziplinärer Expertise essentiell. Die empirischen Wirtschaftswissenschaften können aufgrund ihres Fokus auf Datengenerierung und menschliche Entscheidungen einiges beitragen, aber auch andere empirisch arbeitenden Disziplinen bieten hilfreiche Ansätze.¹⁴ In jeder Strategie zur Förderung

¹¹ Vgl. zum Beispiel Hauke Hohensee (2019): Digitaltechnik in der Medizin: Was bringt smart Technik für die Gesundheit? Der Tagesspiegel, 11. März 2019 (online verfügbar).

¹² Eliza Strickland (2019): IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. IEEE Spectrum, 56(4), 24-31.

¹³ Vgl. Janet M. Currie und W. Bentley MacLeod (2018): Understanding physician decision making: the case of depression. NBER Working Paper Nr. w24955. Kleinberg et al. (2018): a. a. O. Himabindu Lakkaraju et al. (2017): The selective labels problem: evaluating algorithmic predictions in the presence of unobservables. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 275-284.

¹⁴ Susan Athey (2018): The impact of machine learning on economics. In: Ajay K. Agrawal, Joshua Gans und Avi Goldfarb (Hrsg.): The economics of artificial intelligence: an agenda. University of Chicago Press.

Kasten 2

Maschinelles Lernen am Beispiel des *random forest*

Ein *random forest* ist eine Sammlung von Klassifikations- oder Regressionsbäumen. Ein Regressionsbaum generiert den bedingten Erwartungswert $E[Y|X]$ der Ergebnisvariable Y basierend auf einer Sammlung von M beobachtbaren Variablen X . Nimmt die Variable Y nur Werte von 0 und 1 an, so generiert ein Regressionsbaum die bedingte Wahrscheinlichkeit, dass im vorliegenden Beispiel eine bakterielle Ursache vorliegt. In einem vereinfachten Beispiel mit zwei Variablen wird gezeigt, wie der Algorithmus Variable für Variable nach der besten Gruppierung nach X sucht, um die Vorhersagequalität von $E[Y|X]$ zu maximieren (Abbildung 3). Diesen Prozess nennt man Trainieren, welcher anhand von Trainingsdaten stattfindet, die von den Daten abweichen, die für Vorhersagen verwendet werden.

Regressionsbäume haben eine sehr geringe Verzerrung, das heißt einzelne Beobachtungen können sehr gut vorhergesagt werden. Dafür führen sie aber zu Vorhersagefehlern bei Daten, die sich von den Ausprägungen in den Trainingsdaten unterscheiden. Um diese Varianz zu verringern werden in einem *random forest* weitere Schritte durchgeführt:

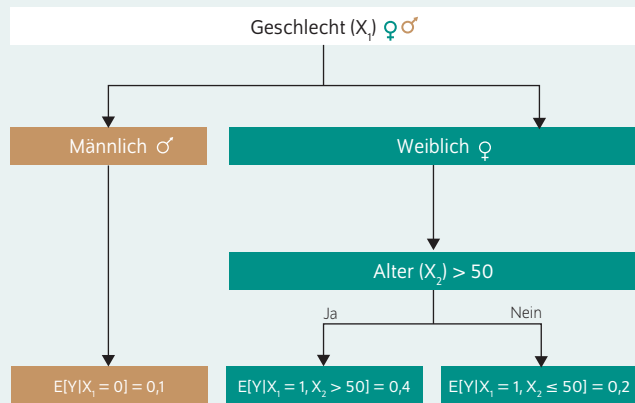
1. Für jeden Baum wird eine randomisierte Auswahl an Beobachtungen aus dem Trainingsdatensatz verwendet.
2. An jedem Knotenpunkt im Baum werden $m < M$ Variablen zufällig gezogen, für welche Gruppierungen gesucht werden, die die Vorhersagequalität maximieren.
3. Die Vorhersage wird für jede Beobachtung als Durchschnitt aller bedingten Erwartungswerte aus allen Bäumen berechnet.

künstlicher Intelligenz ist diese Expertise für erfolgreiche Umsetzungen in vielfältigen Einsatzgebiete über technische Investitionen hinaus von entscheidender Bedeutung.

Die zu bewältigenden, nicht-trivialen Herausforderungen weisen darauf hin, dass Methoden des maschinellen Lernens zwar das Potential für beachtlichen gesellschaftlichen

Abbildung 3

Stilisierte Regressionsbaum



© DIW Berlin 2019

Im vorliegenden Beispiel ist Y gleich eins, wenn eine bakterielle Infektion vorliegt, andernfalls null. Der Baum sucht nach Eigenschaften, die mit diesen Infektionen korrelieren.

Dank der dadurch erzielten hohen Vorhersagequalität gehört der einfach implementierbare und sehr flexible *random forest* zu den führenden Standard-Algorithmen unter den Methoden des maschinellen Lernens.

Fortschritt haben, ihre vollen Auswirkungen aber schrittweise erzielt werden und eine sprunghafte Revolution vermutlich zunächst ausbleibt. Um Potentiale voranzutreiben und gleichzeitig die gesellschaftliche Akzeptanz zu fördern, werden experimentelle Erprobungen in Zusammenarbeit mit fachspezifischen ExpertInnen und begleitenden Evaluierungen unerlässlich sein.

Michael A. Ribers ist wissenschaftlicher Mitarbeiter am DIW Berlin und Postdoktorand an der Universität Kopenhagen | mribers@diw.de

Hannes Ullrich ist wissenschaftlicher Mitarbeiter am DIW Berlin und Associate Professor an der Universität Kopenhagen | hullrich@diw.de

JEL: C10; C55; I11; I18; L38; O38; Q28

Keywords: antibiotic prescribing; prediction policy; machine learning; expert decision-making

This report is also available in an English version as DIW Weekly Report 19/2019:

www.diw.de/diw_weekly



IMPRESSUM



DIW Berlin — Deutsches Institut für Wirtschaftsforschung e.V.

Mohrenstraße 58, 10117 Berlin

www.diw.de

Telefon: +49 30 897 89-0 Fax: -200

86. Jahrgang 7. Mai 2019

Herausgeberinnen und Herausgeber

Prof. Dr. Tomaso Duso; Prof. Marcel Fratzscher, Ph.D.; Prof. Dr. Peter Haan;
Prof. Dr. Claudia Kemfert; Prof. Dr. Alexander Kriwoluzky; Prof. Dr. Stefan Liebig;
Prof. Dr. Lukas Menkhoff; Dr. Claus Michelsen; Prof. Karsten Neuhoff, Ph.D.;
Prof. Dr. Jürgen Schupp; Prof. Dr. C. Katharina Spieß

Chefredaktion

Dr. Gritje Hartmann; Mathilde Richter; Dr. Wolf-Peter Schill

Lektorat

Felix Weinhardt, Ph.D. (erster Bericht); Prof. Dr. Jan Marcus (zweiter Bericht)

Redaktion

Renate Bogdanovic; Dr. Franziska Bremus; Rebecca Buhner;
Claudia Cohnen-Beck; Dr. Daniel Kemptner; Sebastian Kollmann;
Bastian Tittor; Dr. Alexander Zerrahn

Vertrieb

DIW Berlin Leserservice, Postfach 74, 77649 Offenburg

leserservice@diw.de

Telefon: +49 1806 14 00 50 25 (20 Cent pro Anruf)

Gestaltung

Roman Wilhelm, DIW Berlin

Umschlagmotiv

© imageBROKER / Steffen Diemer

Satz

Satz-Rechen-Zentrum Hartmann + Heenemann GmbH & Co. KG, Berlin

Druck

USE gGmbH, Berlin

ISSN 0012-1304; ISSN 1860-8787 (online)

Nachdruck und sonstige Verbreitung – auch auszugsweise – nur mit
Quellenangabe und unter Zusendung eines Belegexemplars an den
Kundenservice des DIW Berlin zulässig (kundenservice@diw.de).

Abonnieren Sie auch unseren DIW- und/oder Wochenbericht-Newsletter
unter www.diw.de/newsletter