

Moritz, Steffen; Wiyneck, Frederik; Wiebels, Johannes

## Article

Prognose der Abgabequote von Einkommensteuererklärungen bei Rentnerinnen und Rentnern: Machbarkeitsstudie zur Beschleunigung der Veröffentlichung von statistischen Ergebnissen mittels Machine Learning

WISTA - Wirtschaft und Statistik

## Provided in Cooperation with:

Statistisches Bundesamt (Destatis), Wiesbaden

*Suggested Citation:* Moritz, Steffen; Wiyneck, Frederik; Wiebels, Johannes (2024) : Prognose der Abgabequote von Einkommensteuererklärungen bei Rentnerinnen und Rentnern: Machbarkeitsstudie zur Beschleunigung der Veröffentlichung von statistischen Ergebnissen mittels Machine Learning, WISTA - Wirtschaft und Statistik, ISSN 1619-2907, Statistisches Bundesamt (Destatis), Wiesbaden, Vol. 76, Iss. 2, pp. 83-96

This Version is available at:

<https://hdl.handle.net/10419/294178>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# PROGNOSE DER ABGABEQUOTE VON EINKOMMENSTEUERERKLÄRUNGEN BEI RENTNERINNEN UND RENTNERN

Machbarkeitsstudie zur Beschleunigung der Veröffentlichung von statistischen Ergebnissen mittels Machine Learning

Steffen Moritz, Frederik Wiyneck, Johannes Wiebels

↳ **Schlüsselwörter:** Machine Learning – Renten – Rentenbezugsmitteilungen – Frühschätzungen – Nowcast – Schätzverfahren

## ZUSAMMENFASSUNG

Jährlich veröffentlicht das Statistische Bundesamt Statistiken über die Besteuerung von Rentnerinnen und Rentnern, wegen langer Abgabe- und Einspruchsfristen für Einkommensteuererklärungen allerdings erst etwa 3,5 Jahre nach Ablauf des betreffenden Statistikjahres. Jedoch liegt ein Teil dieser Einkommensteuerdaten bereits deutlich früher vor. Das Statistische Bundesamt hat in einer Machbarkeitsstudie untersucht, ob Machine-Learning-Modelle geeignet sind, um Daten der Rentenbesteuerung vorherzusagen, bevor sie vollständig vorliegen. Das Ziel ist eine beschleunigte Veröffentlichung dieser Statistiken. Der Artikel erläutert das Vorgehen bei der Suche nach dem besten Schätzalgorithmus und wie die Schätzergebnisse im Hinblick auf Genauigkeit und Robustheit zu evaluieren sind.

↳ **Keywords:** machine learning – pensions – pension benefit notifications – early estimates – nowcast – estimation procedure

## ABSTRACT

*Every year, the Federal Statistical Office of Germany publishes statistics about the taxation of pensioners. Due to long deadlines for the filing of income tax returns and the lodging of objections, 3.5 years can elapse between the end of the relevant statistical year and the publication of the statistics. However, some of these income tax data are available much earlier. The Federal Statistical Office conducted a feasibility study to examine whether machine learning models are suitable for predicting data on the taxation of pensioners before all the data are available. The goal is to enable the swifter publication of the statistics.*

### Steffen Moritz

hat einen Master of Science in Wirtschaftsinformatik und ist als Referent mit Verfahren des maschinellen Lernens und der Imputation im Referat „Künstliche Intelligenz, Big Data“ des Statistischen Bundesamtes befasst. Er beschäftigt sich dabei unter anderem mit den methodischen Fragestellungen beim Einsatz dieser Verfahren und setzt gemeinsam mit den Fachbereichen Anwendungen um.

### Dr. Frederik Wiyneck

ist promovierter Volkswirt und hat an den Universitäten Bayreuth und Erlangen-Nürnberg studiert. Seit 2021 ist er wissenschaftlicher Mitarbeiter im Referat „Lohn- und Einkommensteuer“ des Statistischen Bundesamtes. Dort arbeitet er an einem Projekt zur Frühschätzung der Einkommensteuerstatistiken mittels Machine Learning.

### Johannes Wiebels

ist Diplom-Volkswirt und hat an der Universität Trier studiert. Er ist seit 2021 Referent im Referat „Lohn- und Einkommensteuer“ des Statistischen Bundesamtes. Seine Aufgabenschwerpunkte liegen in der methodischen Weiterentwicklung sowie Erstellung und Verbreitung von Ergebnissen der Statistik der Rentenbezugsmitteilungen.

## 1

### Einleitung

---

Wichtige Qualitätskriterien bei der Statistikproduktion durch die Statistischen Ämter des Bundes und der Länder, die den sogenannten Statistischen Verbund bilden, sind Genauigkeit und Zuverlässigkeit der veröffentlichten Ergebnisse. Informationen stoßen jedoch auf umso mehr Interesse, je schneller sie zur Verfügung stehen. Daher stellt die Aktualität der veröffentlichten Daten ein weiteres, häufig jedoch zu den obigen Kriterien konträres Qualitätsmerkmal für statistische Produkte dar (Statistische Ämter des Bundes und der Länder, 2021). Dieser Artikel beschreibt eine Machbarkeitsstudie, die evaluiert hat, inwiefern Machine Learning (ML) geeignet ist, veröffentlichungsrelevante Informationen zur Rentenbesteuerung vorherzusagen, bevor die zugrunde liegenden Daten vollständig zur Verfügung stehen. Ziel des Einsatzes solcher Machine-Learning-Verfahren ist, bei konstanter Genauigkeit die Aktualität von statistischen Ergebnissen zu steigern.<sup>1</sup>

Die Bevölkerung in Deutschland altert im Zuge des demografischen Wandels und die langfristige Finanzierung des gesetzlichen Rentensystems steht zur Diskussion; daher sind Renten dauerhaft ein Thema von hoher gesellschaftlicher und politischer Relevanz. Das Statistische Bundesamt verarbeitet für die Lohn- und Einkommensteuerstatistik (LESt) auch die Daten der an die Finanzverwaltung übermittelten Rentenbezugsmitteilungen<sup>2</sup> (RBM). Auf dieser Grundlage erfolgt jährlich die Veröffentlichung von Tabellen zur Anzahl der Rentnerinnen und Rentner sowie zu deren ausgezahlten Renten. Wie viele der Rentnerinnen und Rentner eine Einkommensteuererklärung abgegeben haben und wie viel Einkommensteuer sie zahlen, wird 3,5 Jahre später veröffentlicht. Diese Daten resultieren aus dem Anspielen von Angaben aus der Lohn- und Einkommensteuerstatistik, die wegen der teilweise langen Fristen zur Abgabe von Steuererklärungen und der Dauer der Steuerfestsetzung erst später verfügbar sind.

---

1 Das Statistische Bundesamt führt Machbarkeitsstudien durch, um die Statistikproduktion stetig zu optimieren. Beispiele dafür sind Dickopf und andere (2019) zur Aktualitätssteigerung sowie Feuerhake/Dumpert (2016) für den Einsatz von Machine Learning zur Effizienzsteigerung von statistischen Arbeitsabläufen.

2 Zur Definition der Rentenbezugsmitteilungen und wie sie erfolgen müssen siehe § 22a Einkommensteuergesetz.

Die Verknüpfung von Lohn- und Einkommensteuerstatistik und Rentenbezugsmitteilungen erfolgt bisher erst, wenn die Daten zur Lohn- und Einkommensteuerstatistik vollständig vorliegen. Diese Studie prüft, ob früh vorliegende LESt-Daten in Verbindung mit den RBM-Daten geeignet sind, zeitnah vorherzusagen, welche Rentnerinnen und Rentner eine Einkommensteuererklärung abgeben.

Der Aufsatz ist wie folgt gegliedert: Kapitel 2 beschreibt die Datenquellen der Rentenbezugsmitteilungen und der Lohn- und Einkommensteuerstatistik. Kapitel 3 erläutert die Ziele der Schätzungen und wie sie durchgeführt wurden. Nach welchen Kriterien die Eignung der Schätzungen für die Umsetzung bewertet werden, zeigt Kapitel 4, während Kapitel 5 Limitationen bei dieser Bewertung benennt. Kapitel 6 präsentiert die Schätzergebnisse, mit einem Fazit und Handlungsempfehlungen beschließt Kapitel 7 den Beitrag.

## 2

### Datengrundlage

---

Die Studie nutzt als Datenquellen die [Rentenbezugsmitteilungen](#) und die [Lohn- und Einkommensteuerstatistik](#). Die RBM-Daten werden auf Grundlage des Alterseinkünftegesetzes erhoben, welches die Besteuerung von Alterseinkünften – insbesondere den Übergang von einer vorgelagerten zu einer nachgelagerten Besteuerung der gesetzlichen Renten – regelt. Demnach werden die Aufwendungen zur Alterssicherung in der Ansparphase schrittweise steuerfrei gestellt und die Leistungen in der Auszahlungsphase stufenweise steuerlich belastet. Um eine ordnungsgemäße Besteuerung der ausgezahlten Leistungen sicherzustellen, wurde das [Rentenbezugsmitteilungsverfahren](#) implementiert. Dieses Verfahren verpflichtet Institutionen, die steuerpflichtige Renten gemäß § 22 Nr. 1 und 5 Einkommensteuergesetz auszahlen, gesetzlich dazu, vordefinierte Angaben über diese Renten und deren Empfänger an die Finanzverwaltung zu übermitteln. Seit 2015 werden diese Informationen auf Basis des Steuerstatistikgesetzes einmal im Jahr zur Aufbereitung und Veröffentlichung an den Statistischen Verbund weitergeleitet.

Die Statistik der Rentenbezugsmitteilungen basiert auf zwei Datensätzen. Im [Rentendatensatz](#) werden die ausgezahlten Renten einzeln erfasst. Eine Person kann mehrere Renten aus verschiedenen Quellen beziehen. Für jedes Statistikjahr werden die Angaben zu den einzelnen Renten einer Person zusammengefasst. Dieser [Personendatensatz](#), auf dem die veröffentlichten Statistiken größtenteils beruhen, enthält je Person somit eine Beobachtungszeile. Für die Studie standen RBM-Daten der Jahre 2015 bis 2021 zur Verfügung. In dieser Zeit enthielt der Datensatz jährlich zwischen 21 Millionen und 22 Millionen Beobachtungen.

Etwa 3,5 Jahre nach Ende eines RBM-Statistikjahres werden die Personendaten mit den Daten der Lohn- und Einkommensteuerstatistik zusammengeführt. Durch die Verknüpfung wird das Analysepotenzial der RBM-Daten erweitert und damit werden detaillierte Aussagen zum Rentenbesteuerungsverfahren sowie zu weiteren neben Renten erzielten Einkünften ermöglicht. Die LESt-Daten basieren auf den Steuererklärungen und Lohnsteuerbescheinigungen aller in Deutschland einkommensteuerpflichtigen Personen. Beobachtungseinheit sind hierbei die Steuerpflichtigen. Ein Steuerpflichtiger kann je nach Veranlagungsform entweder eine einzelne Person oder ein gemeinsam veranlagtes Ehepaar sein. Für die Verknüpfung mit den Rentenbezugsmitteilungen werden die Daten der steuerpflichtigen Paare auf die Ebene von Einzelpersonen konvertiert. In der Studie wurden für die LESt-Daten wegen methodischer Umstellungen in der Statistik keine Daten vor 2012 berücksichtigt. Die Daten der Jahre nach 2018 waren zum Zeitpunkt der Testrechnungen noch unvollständig und daher für die Studie nicht nutzbar. Von 2012 bis 2018 ist die Anzahl der Personen in der Lohn- und Einkommensteuerstatistik von 52 Millionen auf 57 Millionen angestiegen. Für die meisten Auswertungen (siehe Kapitel 6) wurde eine 1-%-Zufallsstichprobe aller Personen aus den RBM-Daten und deren angespielten LESt-Daten ( $n \approx 210\,000$ ) verwendet.<sup>3</sup>

Der [Zeitverzug](#) zwischen Erstveröffentlichung der RBM-Statistik und Veröffentlichung von Ergebnissen nach Verknüpfung mit den LESt-Daten ist auf die langen Abgabe- und Widerspruchsfristen im steuerlichen Veranlagungsprozess zurückzuführen. Dies hat zur Folge, dass

Informationen aus der Lohn- und Einkommensteuerstatistik für ein Statistikjahr  $t$  erst zum Zeitpunkt  $t+3,5$  vollständig verfügbar sind. Die Daten zu Einkommensteuererklärungen werden in drei Teilmengen an den Statistischen Verbund geliefert: 9, 21 und 33 Monate nach Ablauf eines Statistikjahres.<sup>4</sup> Diese Teilmengen enthalten jeweils die Daten aller bis zum jeweiligen Zeitpunkt festgesetzten Einkommensteuererklärungen des Steuerjahres. Nach 33 Monaten werden außerdem Informationen aus noch offenen Lohnsteuermeldungen geliefert, für die noch keine Steuererklärung abgegeben wurde. Für das Statistikjahr 2018 beispielsweise enthielt die erste Lieferung Daten zu etwa 35% der Steuerpflichtigen, die zweite 29%, die dritte 6% und die der offenen Lohnsteuerfälle 30%. Insbesondere die Informationen der Teilmengen 1 und 2 sind für eine Frühschätzung wertvoll.

### 3

## Ziel der Schätzungen und gewähltes Vorgehen

Die Schätzungen sollen eine frühere Veröffentlichung der RBM-Statistik ermöglichen. Die folgende [Tabelle 1](#) zeigt die angestrebten Frühschätzungen, integriert in die [bestehende Veröffentlichungstabelle](#) der Statistik für den Zeitraum 2015 bis 2021. Die rot dargestellten Werte in den Spalten 3 und 4 liegen zum in der Tabelle angenommenen Veröffentlichungszeitpunkt im Jahr 2022 nicht endgültig vor. Im Idealfall könnten diese jedoch mit einem geeigneten Machine-Learning-Verfahren auf Basis aller zum Schätzzeitpunkt vorliegenden und relevanten RBM- und LESt-Daten geschätzt werden. Eine Ergebnisbereitstellung wäre dann bereits sechs Monate nach Ende des Statistikjahres möglich, statt wie bisher nach 3,5 Jahren. Perspektivisch wären dann auch Veröffentlichungen aufgeschlüsselt nach Altersgruppen, Geschlecht oder Einkommensgruppen prognostizierbar. Nachfolgend soll geprüft werden, ob fehlende Daten mittels Machine Learning genau und robust genug geschätzt werden können, um Frühschätzungen zu veröffentlichen.

<sup>3</sup> Einzelne Tests mit höheren Stichprobenumfängen zeigten keine weitere Verbesserung der Schätzgüte.

<sup>4</sup> Weil die Daten nicht immer exakt nach 33 Monaten vorliegen und die gelieferten Teilmengen final aufbereitet werden, wird die Dauer bis zur vollständigen Datenverfügbarkeit vereinfacht mit 3,5 Jahren angegeben.

**Tabelle 1**

Beispiel für eine Erweiterung der Statistik der Rentenbezugsmitteilungen basierend auf Machine-Learning-Verfahren

	Insgesamt		Davon haben eine Steuererklärung abgegeben	
	Personen	Rentenleistungen	Personen	
	1 000	Mill. EUR	1 000	%
	1	2	3	4
2015	21 288	279 315	10 410	48,9
2016	21 304	289 086	10 588	49,7
2017	21 366	300 191	10 875	50,9
2018	21 553	311 189	11 294	52,4
2019	21 645	327 534	11 515	53,2
2020	21 791	341 052	11 854	54,4
2021	21 918	349 507	12 121	55,3

ML-Algorithmen erlernen, ähnlich wie auch andere statistische Schätzverfahren, die Zusammenhänge zwischen erklärenden und abhängigen Variablen anhand von Beobachtungen, für die sowohl die Werte der erklärenden Variablen als auch der Wert der abhängigen Variable bekannt sind. Diese Fälle liefern den Algorithmen die Informationen, welche Werte die abhängige Variable bei unterschiedlichen Ausprägungen der erklärenden Variablen annimmt. So können die Algorithmen Schätzwerte für Beobachtungen berechnen, bei denen der Wert der abhängigen Variable nicht bekannt ist. Zur Auswahl des besten ML-Algorithmus wird die Anwendungssituation simuliert. Hierzu werden auf einem Teil der zur Verfügung stehenden Daten (Trainingsdaten) ML-Modelle angeleitet. Dann werden mit diesen Modellen Schätzungen für einen anderen Teil der Daten (Testdaten) erstellt. Die realen Werte werden anschließend mit den verschiedenen Schätzungen verglichen. So kann evaluiert werden, wie genau und wie robust die Schätzergebnisse der unterschiedlichen Algorithmen sind.

Fokus der Machbarkeitsstudie ist zunächst, auf Ebene von Einzelpersonen vorherzusagen, welche Rentne-

**Tabelle 2**

Fiktives Beispiel für Beobachtungen im Schätzdatensatz

Steuer-ID	Alter	Geschlecht	Renten <sub>t-1</sub>	Renten <sub>t</sub>	Einkommen <sub>t-1</sub>	Einkommen <sub>t-2</sub>	Einkommensteuererklärung abgegeben	
	Jahre		EUR				t-1	t
10101010101	75	0 (Mann)	12 000	12 000	-	-	0	0
10101010102	66	1 (Frau)	20 000	23 000	33 000	42 000	1	1
10101010103	86	0 (Mann)	12 000	14 000	15 000	17 000	1	0

rinnen und Rentner eine Einkommensteuererklärung abgegeben werden. Bei zufriedenstellenden Ergebnissen können in nächsten Schritten auch Inhalte der Steuererklärung geschätzt werden, um detailliertere Frühschätzungen zu veröffentlichen. Erst muss jedoch bekannt sein, ob eine Person eine solche Erklärung abgibt, bevor es sinnvoll ist, deren Inhalte vorherzusagen. Liefert dieser Analyseschritt keine zufriedenstellenden Ergebnisse, sind weitere Vorhersagen überflüssig. Die abhängige Variable, die zu schätzen ist, ist somit die folgende binäre Variable:

- (1) EStE abgegeben<sub>t</sub> = Rentner/-in hat für Jahr t eine Steuererklärung abgegeben (ja = 1 vs. nein = 0)

Der Output ist ein Prognosewert zwischen 0 und 1, der für manche Algorithmen einer Wahrscheinlichkeit entspricht. In diesem Fall steht zum Beispiel ein Wert von 0,25 für die geschätzte Wahrscheinlichkeit von 25%, dass die betreffende Person eine Steuererklärung abgibt.

Erklärende Variablen sind beispielsweise Alter, Geschlecht, Vorjahreseinkommen, Besteuerungsanteil der Rente oder die Information, ob die betreffende Person vor einem oder vor zwei Jahren eine Steuererklärung abgegeben hat. Diese und weitere Informationen aus vergangenen Rentenbezugsmitteilungen und Einkommensteuererklärungen sind nutzbar, weil die RBM- und LESt-Daten einer Person über ihre Steuer-Identifikationsnummer (Steuer-ID)<sup>5</sup> auch zwischen verschiedenen Jahren verknüpfbar sind. Für die Schätzungen wird entsprechend ein Datensatz erstellt, der Daten aus Rentenbezugsmitteilung und Lohn- und Einkommensteuererklärung des zu schätzenden Jahres sowie aus den Vorjahren beider Quellen enthält.

➤ Tabelle 2 zeigt beispielhaft, wie ein Schätzdatensatz aufgebaut ist. Immer vorhanden ist die Steuer-ID als

<sup>5</sup> Die Steuer-ID ist eine elfstellige Zahl, die jeder steuerzahlenden Person von der Finanzverwaltung zugewiesen wird, sie eindeutig identifiziert und für jede Person lebenslang gleich bleibt.

Identifikator. Weiterhin sind Personenmerkmale wie Alter oder Geschlecht verfügbar. Es gibt Variablen zu den Rentenbezügen, die aus der aktuellen Rentenbezugsmitteilung abgeleitet sind ( $Renten_t$ ), aber auch Vorjahresdaten zu Rentenbezügen ( $Renten_{t-1}$ ). Außerdem gibt es Variablen aus vergangenen Steuerjahren (Einkommen $_{t-1}$ , Einkommen $_{t-2}$ ), sofern für das entsprechende Jahr zum Schätzzeitpunkt bereits eine Steuererklärung abgegeben wurde. Abschließende Variable ist die in den Testszenarien vorhandene, aber zum Zeitpunkt der Schätzung in der Anwendungssituation noch unbekannte abhängige Variable, ob im Jahr  $t$  eine Steuererklärung abgegeben wurde.

Ein vollständig integrierter RBM- und LEST-Datensatz hat – jeweils ableitbare Vorjahresvariablen noch unberücksichtigt – mehr als 1 500 Variablen. Diese große Anzahl an für die Frühschätzung verwendbaren Variablen stellt eine Herausforderung in Bezug auf ihre Handhabbarkeit dar. Deshalb erfolgte eine fachliche Vorauswahl, bei der 365 Variablen als potenziell relevant identifiziert wur-

den. Von diesen wiesen wiederum 279 eine positive Variablenwichtigkeit (siehe Anmerkung in Grafik 1) auf, was darauf hindeutet, dass sie einen positiven Beitrag zur Güte der ML-Modelle leisten können.

➤ Grafik 1 zeigt die 13 wichtigsten erklärenden Variablen. Je höher der Wert für eine Variable, desto mehr würde sich die Schätzgenauigkeit verschlechtern, würde sie nicht in die Schätzungen einbezogen.

## 4

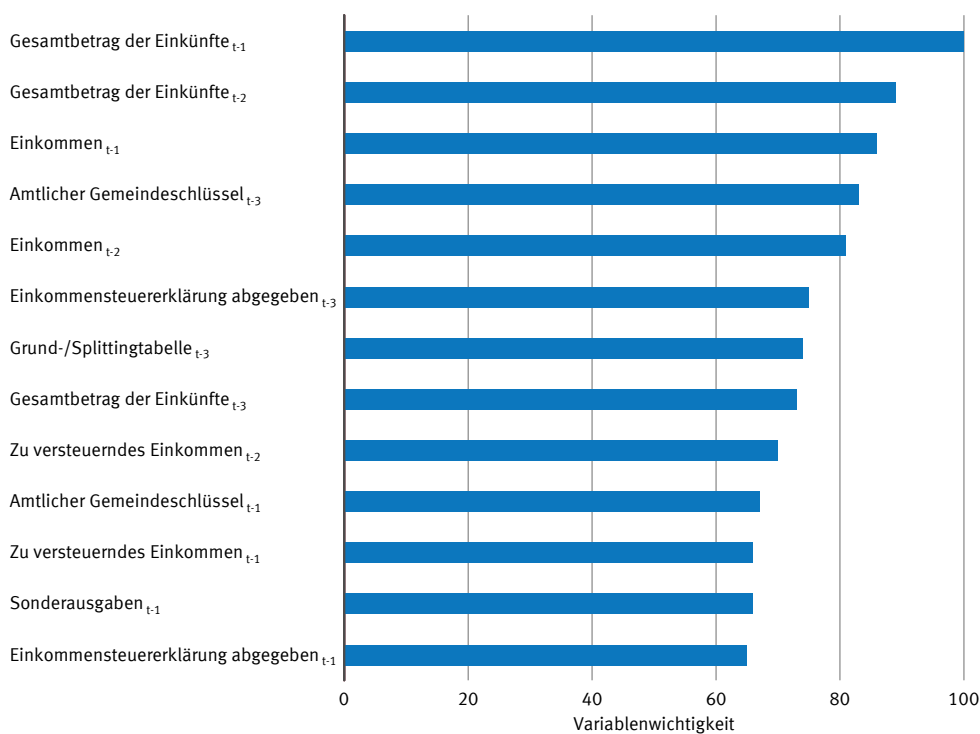
### Bewertungskriterien

Die Veröffentlichung von Frühschätzungen wäre gerechtfertigt, wenn sie ausreichend genau und robust sind.<sup>16</sup> Die **Genauigkeit** einer Schätzung wird danach beurteilt, wie nah sie am tatsächlichen Ergebnis liegt. Um diese

<sup>6</sup> Siehe hierzu auch Saidani und andere (2023).

**Grafik 1**

Wichtigkeit der erklärenden Variablen bei der Vorhersage der abhängigen Variable in %



**Anmerkung:** Die Grafik bezieht sich auf die mittels Permutation Importance berechnete Variablenwichtigkeit eines Random-Forest-Modells (ranger). Ausgangspunkt war eine Schätzung für die Abgabe einer Steuererklärung im Jahr 2018 mit Trainingsdaten aus 2017. Die x-Achse vergleicht den relativen Erklärungsgehalt der jeweiligen Variablen in der Schätzung. Die Zahlen sind nicht absolut interpretierbar.

Aussage treffen zu können, müssen Schätzungen für Beobachtungen durchgeführt werden, bei denen das reale Ergebnis bereits bekannt ist. Nur für diese können Schätzung und Realität verglichen werden.

Zur Bewertung von Schätzverfahren können verschiedene Metriken herangezogen werden. Welche Metrik verwendet wird, ist vom jeweiligen Anwendungsfall abhängig. Als eine Metrik, die zentral und leicht verständlich ist, wurde für diese Studie die **Abgabequote** gewählt, also wie groß der Anteil der Rentnerinnen und Rentner ist, die für ein gegebenes Jahr eine Steuererklärung abgeben werden. Die Abweichung der vorhergesagten von der realen Abgabequote ist eine Metrik für die Genauigkeit. Berechnet wird die vorhergesagte Abgabequote als Mittelwert der geschätzten Abgabewahrscheinlichkeiten über alle Rentnerinnen und Rentner ( $n$ ) hinweg.

$$(2) \text{ Abgabequote}_t = \frac{1}{n} \sum_{i=1}^n P(\text{EStE abgegeben}_t = 1)$$

Als Minimalkriterium für eine ausreichende Genauigkeit der Vorhersage mit ML-Methoden wird deren Genauigkeit mit der einer simplen linearen Extrapolation verglichen. Das genaue Vorgehen wird in Abschnitt 6.2 beschrieben. Wäre eine solche Extrapolation annähernd so genau wie die aufwendigere ML-Schätzung, wäre der Aufwand für Letztere nicht gerechtfertigt.

Generell gilt, dass eine Schätzung umso **robuster** ist, je weniger ihre Genauigkeit in verschiedenen relevanten Szenarien schwankt.<sup>7</sup> Die Abweichung der Abgabequote als einzige Bewertungsmetrik zu nutzen, birgt dabei eine Gefahr. Die vorhergesagte Abgabequote kann sehr nah an der Realität liegen, selbst wenn der Schätzalgorithmus den Zusammenhang zwischen erklärenden Variablen und abhängiger Variable falsch abbildet. Fälschlich zu hoch und zu niedrig geschätzte Wahrscheinlichkeiten können sich bei der Berechnung des Mittelwertes gegenseitig aufheben. Bei Schätzungen für verschiedene Subgruppen oder andere Jahre wären die Ergebnisse wahrscheinlich deutlich ungenauer, weil die hohe Genauigkeit der ursprünglichen Schätzung nur zufällig war.

<sup>7</sup> Robustheit kann dabei mehrere Teilaspekte umfassen (Saidani und andere, 2023). Der Artikel bezieht sich nachfolgend auf die Anwendbarkeit über die Zeit.

Um diese Gefahr zu verringern, wird zusätzlich der **Prediction Error** (Vorhersagefehler) betrachtet. Er definiert den Anteil der falsch klassifizierten an allen geschätzten Beobachtungen. Zur Berechnung werden die geschätzten Abgabewahrscheinlichkeiten für jede Person in eine Klassifizierung umgewandelt. Ist die geschätzte Wahrscheinlichkeit kleiner oder gleich 0,5, lautet die geschätzte Klassifizierung „EStE abgegeben<sub>t</sub> = 0“, liegt sie über 0,5, dann „EStE abgegeben<sub>t</sub> = 1“. Entspricht der geschätzte Wert nicht dem tatsächlichen Wert, ist die Beobachtung falsch klassifiziert.

$$(3) \text{ Prediction Error}_t = \frac{\text{falsch klassifizierte Beobachtungen}_t}{\text{alle Beobachtungen}_t}$$

Hat der Prediction Error zum Beispiel einen Wert von 0,2, so bedeutet dies, dass 20% aller Beobachtungen falsch klassifiziert sind. Eine Schätzung ist umso genauer, je näher der Prediction Error an 0 liegt.

## 5

### Limitationen bei der Bewertung

Um die Robustheit der entwickelten Schätzverfahren einordnen zu können, ist **Backtesting** notwendig – die Anwendung der Verfahren auf historische Daten, für welche die Ergebnisse bereits bekannt sind. Je mehr Perioden der Vergangenheit für das Backtesting zur Verfügung stehen und je diverser diese Perioden sind, desto besser kann die Robustheit eingeschätzt werden. Stünde für das Backtesting nur ein Jahr, also eine Periode, zur Verfügung, könnte auch nur für dieses eine Jahr die Vorhersagegenauigkeit bestimmt werden. Sollte sie extrem hoch sein, könnte dies ein Zufall oder das Ergebnis von Overfitting sein. **Overfitting** bedeutet, dass eine zu große Anpassung des Schätzalgorithmus an die Trainingsdaten erfolgt. Es wäre also nicht klar, ob eine Vorhersage für das nächste Jahr eine ähnliche Vorhersagegenauigkeit bietet.

Liegen dagegen Schätzungen für beispielsweise fünf Jahre vor und die Genauigkeit wäre für jedes Jahr ähnlich hoch, können die Vorhersagen als deutlich verlässlicher angesehen werden. Wie verlässlich, das hängt auch davon ab, wie divers diese fünf Jahre verlaufen sind. Bei dieser Studie stellt dieser Aspekt eine Limitation für die Aussagekraft der Ergebnisse hinsichtlich ihrer Robust-

## Prognose der Abgabequote von Einkommensteuererklärungen bei Rentnerinnen und Rentnern

heit dar. Für das Backtesting stehen die Jahre 2015 bis 2018 zur Verfügung. Diese verliefen wirtschaftlich relativ gleichmäßig. Erst im Jahr 2020 legte die Corona-Pandemie die Wirtschaft lahm und im Jahr 2022 begann der Ukrainekrieg. Wie sich diese Ereignisse auf die abhängige Variable ausgewirkt haben, ist unklar, da zum Zeitpunkt der Studie nicht alle LEST-Daten dieser Jahre vorlagen.

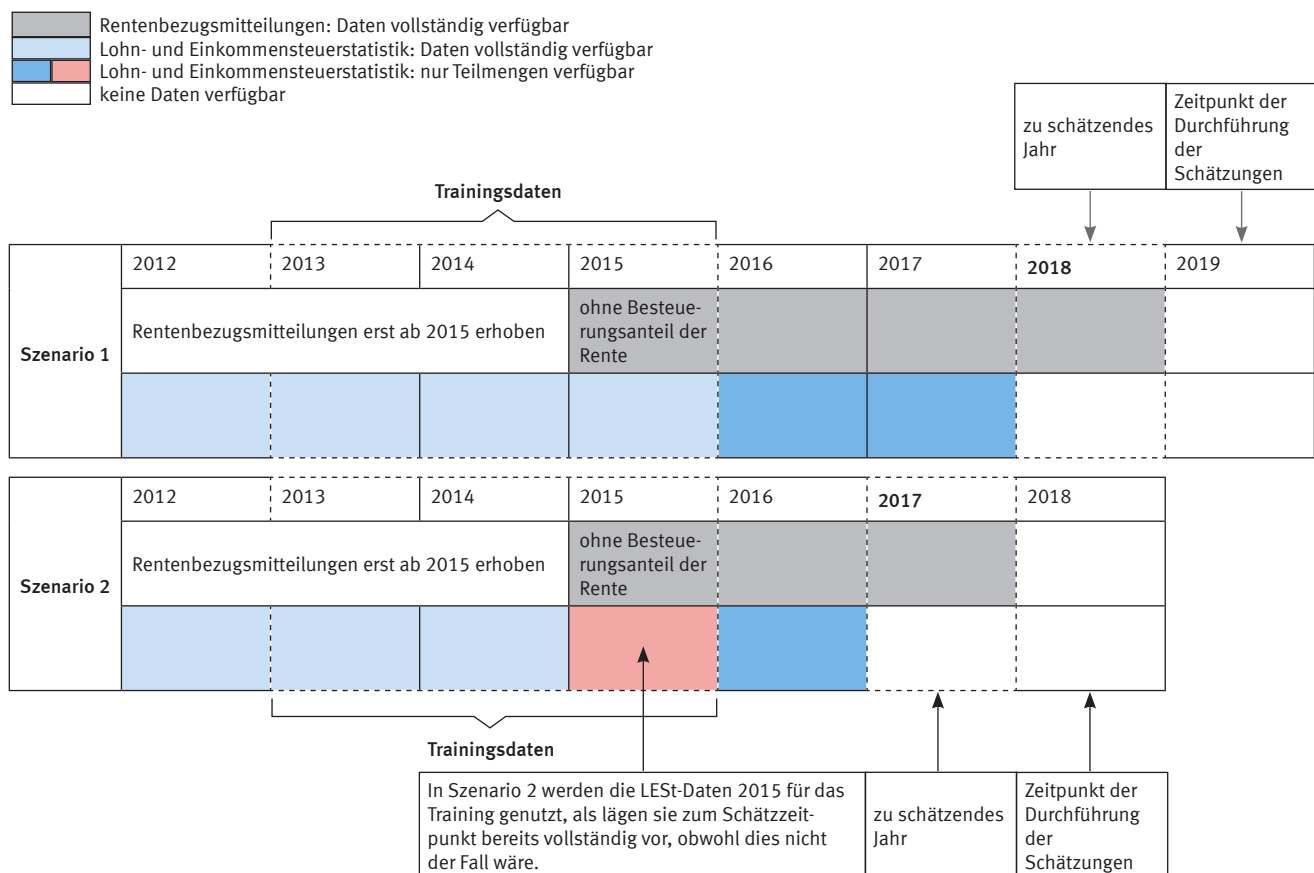
Eine weitere Limitation beim Backtesting ergibt sich aus der Kombination zweier Aspekte: erstens die begrenzte Anzahl von Jahren, für die sowohl RBM- als auch LEST-Daten zur Verfügung stehen, sowie zweitens der zeitliche Verzug von 3,5 Jahren, bis LEST-Daten vollständig vorliegen. Diese Kombination ist problematisch, weil Testrechnungen umso aussagekräftiger sind, je realistischer sie die spätere Anwendungssituation abbilden. Im Rahmen der Studie konnte jedoch keine Schätzung durchgeführt werden, die exakt der Anwendungssituation entspricht,

sondern lediglich Schätzungen, die entweder zu optimistische oder zu pessimistische Genauigkeitsprognosen liefern. [➤ Grafik 2](#) visualisiert die Datenverfügbarkeit in zwei Szenarien.

**Szenario 1** skizziert die Datenverfügbarkeit für eine Schätzung für das Jahr 2018. Zum Zeitpunkt der Studie war 2018 das aktuellste Jahr, für welches die realen Werte der abhängigen Variable bekannt sind. Im Anwendungsfall würde die Schätzung in der ersten Jahreshälfte 2019 stattfinden. Algorithmen benötigen Trainingsdaten, also Daten, für die Werte sowohl der abhängigen als auch der erklärenden Variablen bekannt sind. Zum Prognosezeitpunkt bezögen sich aufgrund des 3,5-jährigen Zeitverzugs die aktuellsten Daten, für die der Wert der abhängigen Variable bekannt ist, auf das Jahr 2015. Zwar lägen bereits die ersten zwei Teilmengen der LEST-Daten aus 2016 und die erste aus 2017 vor. Doch wäre für beide Jahre jeweils nicht sicher, ob Personen,

### Grafik 2

Skizzierung der Datenverfügbarkeit in verschiedenen Szenarien





die noch keine Steuererklärung abgegeben haben, dies noch tun werden. Im Jahr 2015 fehlt in den RBM-Daten die wichtige erklärende Variable Besteuerungsanteil der Rente, die zum Zeitpunkt der Studie erst ab 2016 vorhanden war. Außerdem können keine historischen Variablen aus den Vorjahresdaten generiert werden. Es fehlen somit mehrere Faktoren, von denen zu erwarten ist, dass sie die Schätzungen späterer Jahre genauer machen. Daher entspricht das Szenario nicht der späteren Anwendungssituation und die Prognose ist bezüglich der Genauigkeit tendenziell zu pessimistisch.

**Szenario 2** skizziert die Datenverfügbarkeit für eine Schätzung für das Jahr 2017. Diese Schätzung ist noch ein wenig weiter vom Anwendungsfall entfernt. Zum Prognosezeitpunkt im Jahr 2018 wären die aktuellsten vollständigen LEST-Daten aus dem Jahr 2014. Für dieses Jahr gäbe es jedoch keine RBM-Daten. Da die abhängige Variable nur als Kombination aus RBM- und LEST-Daten erstellt werden kann, lässt sich in einem realistischen Szenario kein Schätzwert für 2017 berechnen. Die nächstbeste Lösung ist, so zu tun, als lägen zum Schätzzeitpunkt bereits die vollständigen LEST-Daten aus dem Jahr 2015 vor. Da die Trainingsdaten in diesem Szenario aber zeitlich näher an den zu schätzenden Daten liegen, als in der Realität möglich, ist die Prognose bezüglich der geschätzten Genauigkeit tendenziell zu optimistisch.

Zu optimistisch wäre das Szenario deshalb, weil es die Möglichkeit eines **Concept Drift** unterschätzt. Concept Drift bedeutet, dass der Zusammenhang zwischen abhängiger Variable und erklärenden Variablen über die Zeit nicht konstant bleibt. Wenn zum Beispiel der Grundfreibetrag ansteigt, entfällt dadurch für manche Personen die Pflicht, eine Einkommensteuererklärung abzugeben. Solche Steuerrechtsänderungen erfolgen jährlich. Je länger der zeitliche Abstand zwischen Trainingsdaten und vorhergesagtem Jahr, desto mehr solcher Unterschiede häufen sich an und desto schlechter bilden die Trainingsdaten die Verhältnisse im geschätzten Jahr ab.

Für die Jahre 2016 und 2015 sind Trainingsdaten aus 2015 ebenfalls die beste Annäherung an ein realistisches Schätzszenario. Hierbei ist die Unterschätzung des Concept Drift noch gravierender, weil Trainingsdaten und geschätztes Jahr noch enger zusammenliegen. Trotzdem wurden die Schätzungen für 2015, 2016 und 2017 für diese Studie durchgeführt, da die Ergebnisse

eine bessere Einschätzung von Genauigkeit und Robustheit zulassen, als läge nur eine Schätzung für das Jahr 2018 vor. Außerdem wurden für 2018 Schätzungen mit Trainingsdaten aus 2016, 2017 und 2018 durchgeführt, ebenso Schätzungen für 2017 und 2016 mit Trainingsdaten, die nicht jünger sind als das jeweilige Jahr. Diese noch etwas unrealistischeren Szenarien dienen dazu zu untersuchen, inwieweit Concept Drift tatsächlich ein Problem ist. Sollte sich zeigen, dass die Ergebnisse je Jahr immer sehr ähnlich genau sind, egal wie alt oder jung die Trainingsdaten sind, würde dies auf geringen Concept Drift schließen lassen.

---

## 6

---

### Ergebnisse

---

---

#### 6.1 Modellvergleich

---

Um den am besten geeigneten Algorithmus für die Vorhersage zu finden, wurden verschiedene Vorhersagemodelle getestet und verglichen. Die Experimente wurden in der Statistiksoftware R mithilfe des ML-Frameworks mlr3 (Lang und andere, 2019) durchgeführt. [↘ Übersicht 1](#) fasst die in der Studie verglichenen Algorithmen zusammen.

Mit diesen ML-Algorithmen wurden verschiedene Schätzszenarien umgesetzt und Genauigkeit und Robustheit der Ergebnisse miteinander verglichen.

[↘ Grafik 3](#) zeigt exemplarisch einen Performance-Vergleich der angewandten Algorithmen für das Schätzszenario 2018 mit Trainingsdaten aus 2017. Zur Erstellung der Modelle wurden für jeden Algorithmus zehnmal zufällig 20 000 Beobachtungen aus 2017 gezogen, auf denen der Algorithmus dann trainiert wurde. Tests haben ergeben, dass selbst eine deutliche Vergrößerung des Trainingsdatensatzes keine genaueren Schätzergebnisse liefert. Die zehn Modelle wurden jeweils auf die gesamten Beobachtungen aus der 1-%-Stichprobe des zu schätzenden Jahres (in diesem Fall 2018) angewendet und der entsprechende Prediction Error berechnet. Die Grafik zeigt einen Boxplot der berechneten Prediction Errors für die zehn trainierten Modelle je Algorithmus.

# Prognose der Abgabequote von Einkommensteuererklärungen bei Rentnerinnen und Rentnern

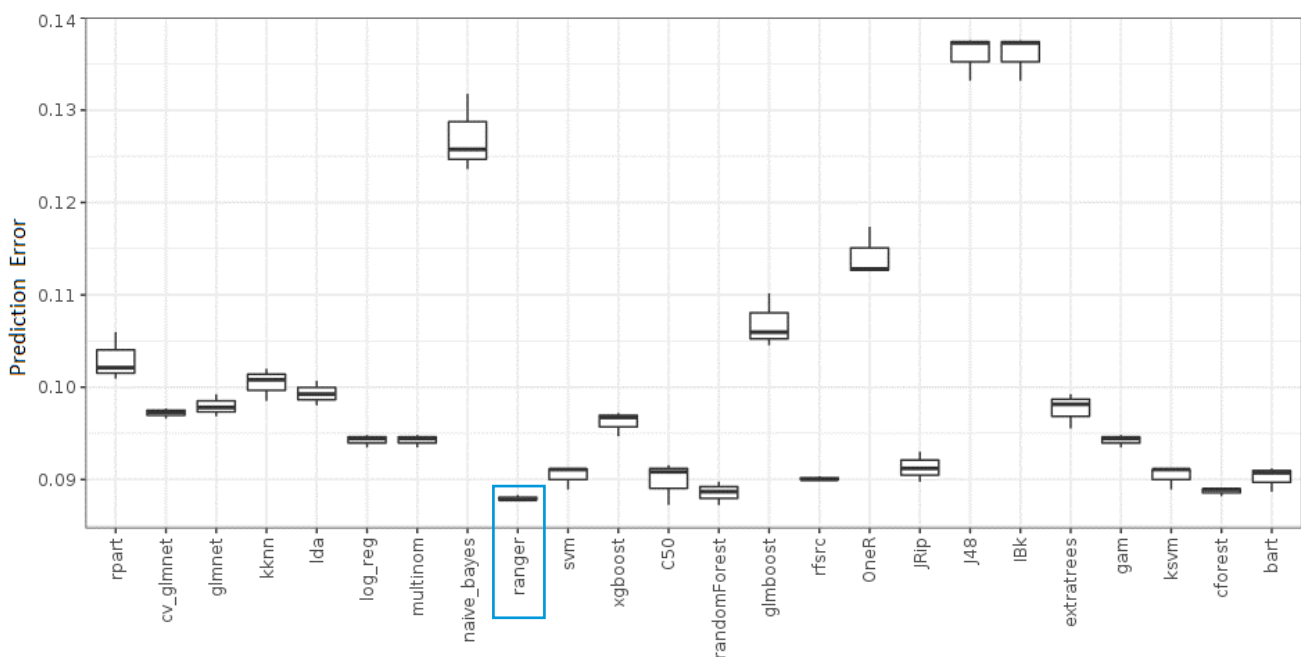
## Übersicht 1

### Übersicht über die verwendeten Machine-Learning-Algorithmen

Einordnung Machine-Learning-Verfahren	Name des Algorithmus	Kurzbezeichnung	R-Paket
Einfache Modelle	Naive Bayes	naive_bayes	e1071 (Meyer und andere, 2023)
	Simple Rule Learner	oneR	RWeka (Hornik und andere, 2009)
	Lineare Diskriminanzanalyse	lda	MASS (Venables/Ripley, 2002)
	Logistische Regression	log_reg	stats (R Core Team, 2023)
Entscheidungsbäume	C5.0 Entscheidungsbäume	c50	C50 (Kuhn/Quinlan, 2023)
	Recursive Partitioning and Regression Trees	rpart	rpart (Therneau/Atkinson, 2022)
Baumbasierte Ensembles	Fast Unified Random Forests	rfsrc	randomForestSRC (Ishwaran und andere, 2008)
	Ranger	ranger	ranger (Wright/Ziegler, 2017)
	Random Forest	randomForest	randomForest (Liaw/Wiener, 2002)
	RF based on conditional inference trees	cForest	partykit (Hothorn/Zeileis, 2015)
	Extremely Randomized Trees	extratrees	extraTrees (Simm und andere, 2014)
	Bayesian Additive Regression Trees	bart	dbarts (Dorie und andere, 2023)
	Extreme Gradient Boosting	xgboost	xgboost (Chen und andere, 2023)
k-Nearest-Neighbor	Weighted k-Nearest Neighbors	kknn	kknn (Schliep/Hechenbichler, 2016)
	k-nearest Neighbor	lbk	RWeka (Hornik und andere, 2009)
Support Vector Machines	Support Vector Machines	svm	e1071 (Meyer und andere, 2023)
	Support Vector Machines	ksvm	kernlab (Karatzoglou und andere, 2004)
Weitere	Multinomial log-linear models via NN	multinom	nnet (Venables/Ripley, 2002)
	Repeated Incremental Pruning	jrip	RWeka (Hornik und andere, 2009)
	Auto-tuned, Penalized Logistic Regression	cv_glmnet	glmnet (Friedman und andere, 2010)

## Grafik 3

### Vergleich der angewendeten Machine-Learning-Algorithmen



**Tabelle 3**  
Ergebnisse

Vorhergesagtes Jahr	Jahr der Trainingsdaten	Vorhandene RBM-Vorjahresvariablen	Prediction Error	Schätzung Abgabequote	Reale Abgabequote	Absolute Abweichung
			%			Prozentpunkte
2018	2018	2 Jahre	12,11	52,37	52,44	0,1
2018	2017	2 Jahre	12,50	53,72	52,44	1,3
2018	2016	1 Jahr	14,98	52,59	52,44	0,2
2018	2015	0 Jahre	13,93	53,43	52,44	1,0
2017	2017	2 Jahre	13,91	50,99	50,90	0,1
2017	2016	1 Jahr	15,07	51,13	50,90	0,2
2017	2015	0 Jahre	14,76	51,54	50,90	0,6
2016	2016	1 Jahr	15,34	49,80	49,73	0,1
2016	2015	0 Jahre	15,46	50,04	49,73	0,3
2015	2015	0 Jahre	15,88	48,74	48,95	0,2

Anmerkung: Die absoluten Abweichungen aus den Schätzungen mit „ranger“ sind umso stärker blau eingefärbt, je näher sie an 0 liegen, und umso stärker rot, je mehr sie von 0 abweichen.

Im abgebildeten Szenario ebenso wie in der Mehrzahl der getesteten Szenarien hatte die **Random-Forest-Implementation „ranger“** den durchschnittlich kleinsten Prediction Error mit sehr geringer Varianz. Die meisten ML-Methoden haben mehrere Stellschrauben, sogenannte **Hyperparameter**, an denen gedreht werden kann, um die Vorhersagegenauigkeit zu optimieren. Allgemein sind dies Einstellungen, die Einfluss auf die exakte Modellbildung nehmen. Für das beste Modell im Vergleich „ranger“ wurde anschließend im Zuge der Studie ein ausführliches Hyperparametertuning betrieben und detaillierte Auswertungen erstellt.<sup>8</sup>

## 6.2 Schätzergebnisse basierend auf dem „ranger“-Algorithmus

Für jedes vorhergesagte Jahr wurden je eine Schätzung mit Trainingsdaten aus dem gleichen und den älteren Jahren durchgeführt. Da der RBM-Datensatz erst ab 2015 zur Verfügung steht, fehlen umso mehr RBM-Vorjahresvariablen, je älter der Trainingsdatensatz ist. [↪ Tabelle 3](#)

Die Ergebnisse zeigen, dass der Prediction Error bei größerem Abstand zwischen vorhergesagtem Jahr und dem Jahr der Trainingsdaten nur geringfügig wächst, was auf

8 Die besten Ergebnisse wurden mit den Parametereinstellungen `mtry = 18`, `min.node.size = 50`, `num.trees = 1600`, `splitrule = „extrarules“`, `replace = TRUE` erzielt.

moderaten Concept Drift schließen lässt. Dies ist ein gutes Zeichen dafür, dass auch die Szenarien, deren Zeitversatz nicht realistisch ist, zu einem gewissen Grad zur Beurteilung von Genauigkeit und Robustheit nutzbar sind.

Um einzuordnen, ob die Ergebnisse genau „genug“ sind, um eine dauerhafte Implementierung der ML-Schätzungen zu rechtfertigen, wurde für jedes Schätzszenario die geschätzte mit der realen Abgabequote verglichen. Die absolute Abweichung wurde in Prozentpunkten ermittelt (siehe Tabelle 3). Im Folgenden werden diese Abweichungen mit solchen verglichen, die mit einer sehr einfachen Methode generiert werden. Die ML-Schätzungen sollten im Idealfall sowohl genauer als auch robuster sein als die einfachen Vergleichsschätzungen.

Die simple Vergleichsschätzung basiert darauf, die letzte bekannte Differenz in Prozentpunkten zum Vorjahr der Abgabequote fortzuschreiben.<sup>9</sup> Im zu schätzenden Jahr  $t$  ist die letzte bekannte Veränderung die Differenz zwischen den drei ( $t-3$ ) und vier Jahre ( $t-4$ ) zurückliegenden Abgabequoten. Für alle aktuelleren Jahre sind die Abgabequoten noch nicht bekannt. Die Vergleichsschätzung (4) wird wie unten dargestellt berechnet.

9 Diese Methode wurde aufgrund ihrer einfachen Erklärbarkeit ausgewählt. Auch andere simple Schätzungen, basierend auf leicht abgewandelten Versionen der hier beschriebenen Formel, wurden geprüft, lieferten aber keine besseren Ergebnisse.

### Berechnung der Vergleichsschätzung

$$(4) \text{ Abgabequote}_t = \text{Abgabequote}_{t-3} + 3 \times (\text{Abgabequote}_{t-3} - \text{Abgabequote}_{t-4})$$

**Tabelle 4**

Vergleich „ranger“-Schätzung und simple Methode

Einkommensteuererklärung- Abgabequote real in %	2017		2018	
	Vorhersage in %	Absolute Abweichung in Prozentpunkten	Vorhersage in %	Absolute Abweichung in Prozentpunkten
	50,90		52,40	
Vergleichsschätzung	50,5	0,39	51,3	1,07
Schätzungen mit ranger:				
Trainingsdaten 2015	51,54	0,65	53,43	0,99
Trainingsdaten 2016	51,13	0,23	52,59	0,16
Trainingsdaten 2017	50,99	0,10	53,72	1,28
Trainingsdaten 2018			52,37	0,07

Rot unterlegt: Schätzungen, die ungenauer sind als die Vergleichsschätzung.

Bei der Umsetzung dieser simplen Methode ergibt sich erneut das Problem, dass erst ab 2015 RBM-Daten zur Verfügung stehen. Daher ist die früheste bekannte Differenz der Abgabequoten die zwischen 2015 und 2016. Deshalb wird für die Vergleichsschätzung diese Differenz von 0,8 Prozentpunkten für die Folgejahre fortgeschrieben.

↘ **Tabelle 4** stellt die Ergebnisse der Vergleichsschätzung denen der ML-Schätzungen gegenüber. Die mittels „ranger“ geschätzten Abgabequoten sind deutlich näher an der Realität als die simplen Prognosen.

Soweit eine Einschätzung möglich ist, fällt das Fazit des Vergleichs eher positiv aus. Es gibt zwar zwei Schätzungen, die ungenauer sind als die Vergleichsschätzung, aber die meisten sind deutlich genauer. Der Vergleich zeigt erneut die Limitationen aufgrund der mangelnden Verfügbarkeit historischer Daten auf. Selbst diese einfache Vergleichsschätzung kann nur unter unrealistischen Verhältnissen und für zwei zu schätzende Jahre durchgeführt werden.

## 7

### Fazit und Ausblick


Für diese Studie wurde untersucht, ob es möglich und sinnvoll ist, einen ML-Algorithmus zu trainieren, der mit ausreichender Genauigkeit vorhersagen kann, ob Rentnerinnen und Rentner eine Einkommensteuererklärung für das abgelaufene Steuerjahr abgeben werden. Dazu wurden verschiedene ML-Methoden auf historische

Daten zu Rentenbezugsmitteilungen und Lohn- und Einkommensteuerstatistik angewendet und die geeignetste – „ranger“ – wurde auf ihre Vorhersagegenauigkeit unter verschiedenen Umständen getestet. Zu klären war die Frage, ob die Ergebnisse ausreichend genau und robust sind, um die Vorhersage jährlich durchzuführen und die Ergebnisse in die Veröffentlichungen des Statistischen Bundesamtes zu den Rentenbezugsmitteilungen einzubeziehen. Der Nutzen hieraus wäre eine schnellere Bereitstellung nachgefragter Informationen, die bereits sechs Monate statt wie bisher 3,5 Jahre nach Ablauf eines Steuerjahres zur Verfügung stünden. Dem entgegen steht die Gefahr, dass künftige Vorhersagen in nicht akzeptablem Ausmaß von der Realität abweichen.

Grundsätzlich stimmen die Ergebnisse der durchgeführten Tests optimistisch, dass eine ausreichend genaue Vorhersage der Abgabe von Steuererklärungen durch Rentnerinnen und Rentner möglich ist. Die absolute Abweichung der vorhergesagten von der tatsächlichen Abgabequote liegt zwischen 0,1 und 1,3 Prozentpunkten. Eine weitere positive Erkenntnis der Tests ist, dass für die untersuchten Jahre, 2015 bis 2018, nur leichter Concept Drift erkennbar ist. Bei einer konkreten Implementierung müssten später weitere für die Veröffentlichung relevante Aspekte berücksichtigt werden. Dazu gehört beispielsweise, wie prognostizierte Ergebnisse für die Nutzerinnen und Nutzer deutlich gekennzeichnet und transparent dokumentiert werden können. Ebenso ist zu klären, welche Prognosewerte in welchem Zyklus aktualisiert werden sollten.

Trotz der positiven Aspekte gibt es zwei Argumente, die gegen eine Umsetzung zum derzeitigen Zeitpunkt sprechen. Erstens basieren die bisherigen Schätzungen

lediglich auf vier Vergleichsjahren, die nur eingeschränkt fürs Backtesting verwertbar sind. Um Schwankungsbreite und Unsicherheiten der Ergebnisse auf aggregierter Ebene abschließend zu beurteilen, wären weitere Jahre notwendig. Die genaue Zeitspanne ist momentan nicht mit Sicherheit feststellbar. Zu erwarten ist, dass eine ausreichend robuste Einschätzung nach fünf weiteren Jahren möglich sein wird. Zweitens haben sich die wirtschaftlichen Bedingungen ab dem Jahr 2020 wegen der Corona-Pandemie und dem Ukrainekrieg drastisch gegenüber den Vorjahren verändert. Während der COVID-19-Pandemie wurden außerdem die Abgabefristen für Steuererklärungen verlängert. Ob die Zusammenhänge, die der Algorithmus aus den Jahren vor diesen Krisen gelernt hat, auch auf die Jahre danach übertragbar sind, lässt sich erst dann endgültig einschätzen, wenn die Daten dieser Jahre vorliegen.

Hierfür müssten die Tests in einigen Jahren mit den zusätzlichen Daten wiederholt werden. Frühestens sollte dies geschehen, nachdem die Daten der Jahre vorliegen, die durch Pandemie und Ukrainekrieg beeinflusst wurden. Sollten die Schätzergebnisse gegenüber diesen Krisen robust nah an der Realität liegen, stünde aus methodischer Sicht einer früheren Veröffentlichung der RBM-Statistiken nichts entgegen. 

## LITERATURVERZEICHNIS

---

Chen, Tianqi/He, Tong/Benesty, Michael/Khotilovich, Vadim/Tang, Yuan/Cho, Hyunsu/Chen, Kailong/Mitchell, Rory/Cano, Ignacio/Zhou, Tianyi/Li, Mu/Xie, Junyuan/Lin, Min/Geng, Yifeng/Li, Yutian/Yuan, Jiaming. *xgboost: Extreme Gradient Boosting*. 2023. [Zugriff am 22. Februar 2024]. Verfügbar unter: [cran.r-project.org](https://cran.r-project.org)

Dickopf, Xaver/Janz, Christian/Mucha, Tanja. *Vom BIP-Flash zum BIP-Nowcast: Erste Ergebnisse einer Machbarkeitsstudie zur weiteren Beschleunigung der BIP-Schnellschätzung*. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2019, Seite 47 ff.

Dorie, Vincent/Chipman, Hugh/McCulloch, Robert. *dbarts: Discrete Bayesian Additive Regression Trees Sampler*. 2023. [Zugriff am 22. Februar 2024]. Verfügbar unter: [cran.r-project.org](https://cran.r-project.org)

Feuerhake, Jörg/Dumpert, Florian. *Erkennung nicht relevanter Unternehmen in den Handwerksstatistiken*. In: WISTA Wirtschaft und Statistik. Ausgabe 2/2016, Seite 79 ff.

Friedman, Jerome H./Hastie, Trevor/Tibshirani, Rob. *Regularization Paths for Generalized Linear Models via Coordinate Descent*. In: Journal of Statistical Software. Band 33. Ausgabe 1/2010, Seite 1 ff. DOI: [10.18637/jss.v033.i01](https://doi.org/10.18637/jss.v033.i01)

Hornik, Kurt/Buchta, Christian/Zeileis, Achim. *Open-source Machine Learning: R Meets Weka*. In: Computational Statistics. Jahrgang 24. Ausgabe 2/2009, Seite 225 ff.

Hothorn, Torsten/Zeileis, Achim. *partykit: A Modular Toolkit for Recursive Partytioning in R*. In: Journal of Machine Learning Research. Jahrgang 16. 2015. Seite 3905 ff. [Zugriff am 22. Februar 2024]. Verfügbar unter: [jmlr.org](https://jmlr.org)

Ishwaran, Hemant/Kogalur, Udaya B./Blackstone, Eugene H./Lauer, Michael S. *Random survival forests*. In: Annals of Applied Statistics. Band 2. Ausgabe 3/2008, Seite 841 ff. [Zugriff am 22. Februar 2024]. Verfügbar unter: [arxiv.org](https://arxiv.org)

Karatzoglou, Alexandros/Smola, Alexandros/Hornik, Kurt/Zeileis, Achim. *kernlab – An S4 Package for Kernel Methods in R*. In: Journal of Statistical Software. Band 11. Ausgabe 9/2004, Seite 1 ff. [Zugriff am 22. Februar 2024]. Verfügbar unter: DOI: [10.18637/jss.v011.i09](https://doi.org/10.18637/jss.v011.i09)

Kuhn, Max/Weston, Steve/Culp, Marc/Coulter, Nathan/Quinlan, Ross. *C50: C5.0 Decision Trees and Rule-Based Models*. 2023. [Zugriff am 22. Februar 2024]. Verfügbar unter: [cran.r-project.org](https://cran.r-project.org)

Lang, Michel/Binder, Martin/Richter, Jakob/Schratz, Patrick/Pfisterer, Florian/Coors, Stefan/Au, Quay/Casalicchio, Giuseppe/Kotthoff, Lars/Bischl, Bernd. *mlr3: A modern object-oriented machine learning framework in R*. In: The Journal of Open Source Software. 2019. DOI: [10.21105/joss.01903](https://doi.org/10.21105/joss.01903)

Liaw, Andy/Wiener, Matthew. *Classification and Regression by randomForest*. In: R News. Band 2. Ausgabe 3/2002, Seite 18 ff. [Zugriff am 22. Februar 2024]. Verfügbar unter: [cran.r-project.org](https://cran.r-project.org)

## LITERATURVERZEICHNIS

---

Meyer, David/Dimitriadou, Evgenia/Hornik, Kurt/Weingessel, Andreas/Leisch, Friedrich. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. [Zugriff am 22. Februar 2024]. Verfügbar unter: [cran.r-project.org](https://cran.r-project.org)

Saidani, Younes/Dumpert, Florian/Borgs, Christian/Brand, Alexander/Nickl, Andreas/Rittmann, Alexandra/Rohde, Johannes/Salwiczek, Christian/Storfinger, Nina/Straub, Selina. *Qualitätsdimensionen maschinellen Lernens in der amtlichen Statistik*. In: AStA Wirtschafts- und Sozialstatistisches Archiv. Band 17. Ausgabe 3–4/2023, Seite 253 ff. DOI: [10.1007/s11943-023-00329-7](https://doi.org/10.1007/s11943-023-00329-7)

Schliep, Klaus/Hechenbichler, Klaus. *kkn: Weighted k-Nearest Neighbors*. 2016. [Zugriff am 22. Februar 2024]. Verfügbar unter: [cran.r-project.org](https://cran.r-project.org)

Simm, Jaak/de Abril, Ildelfons Magrans/Sugiyama, Masashi. *Tree-Based Ensemble Multi-Task Learning Method for Classification and Regression*. In: IEICE TRANSACTIONS on Information and Systems. Band 97. Ausgabe 6/2014, Seite 1677 ff.

Statistische Ämter des Bundes und der Länder. *Qualitätshandbuch der Statistischen Ämter des Bundes und der Länder*. Version 1.21. Wiesbaden 2021. [Zugriff am 22. Februar 2024]. Verfügbar unter: [www.destatis.de](https://www.destatis.de)

Therneau, Terry/Atkinson, Beth. *rpart: Recursive Partitioning and Regression Trees*. 2022. [Zugriff am 22. Februar 2024]. Verfügbar unter: [cran.r-project.org](https://cran.r-project.org)

Venables, W. N./Ripley, B. D. *Modern Applied Statistics with S (Fourth Edition)*. 2002. DOI: [10.1007/978-0-387-21706-2](https://doi.org/10.1007/978-0-387-21706-2)

Wright, Marvin N./Ziegler, Andreas. *ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R*. In: Journal of Statistical Software. Band 77. Ausgabe 1/2017, Seite 1 ff. DOI: [10.18637/jss.v077.i01](https://doi.org/10.18637/jss.v077.i01)

## RECHTSGRUNDLAGEN

---

Einkommensteuergesetz (EStG) in der Fassung der Bekanntmachung vom 8. Oktober 2009 (BGBl. I Seite 3366, 3862), das zuletzt durch Artikel 20 des Gesetzes vom 22. Dezember 2023 (BGBl. I Nr. 411) geändert worden ist.

Gesetz über Steuerstatistiken (StStatG) vom 11. Oktober 1995 (BGBl. I Seite 1250, 1409), das zuletzt durch Artikel 33 des Gesetzes vom 16. Dezember 2022 (BGBl. I Seite 2294) geändert worden ist.

Gesetz zur Neuordnung der einkommensteuerrechtlichen Behandlung von Altersvorsorgeaufwendungen und Altersbezügen (Alterseinkünftegesetz – AltEinkG) vom 5. Juli 2004 (BGBl. I Seite 1427), das zuletzt durch Artikel 117 der Verordnung vom 31. Oktober 2006 (BGBl. I Seite 2407) geändert worden ist.

**Herausgeber**  
Statistisches Bundesamt (Destatis), Wiesbaden

---

**Schriftleitung**  
Dr. Daniel Vorgrimler  
Redaktion: Ellen Römer

---

**Ihr Kontakt zu uns**  
[www.destatis.de/kontakt](http://www.destatis.de/kontakt)

---

**Erscheinungsfolge**  
zweimonatlich, erschienen im April 2024  
Ältere Ausgaben finden Sie unter [www.destatis.de](http://www.destatis.de) sowie in der [Statistischen Bibliothek](#).

---

Artikelnummer: 1010200-24002-4, ISSN 1619-2907

---

© Statistisches Bundesamt (Destatis), 2024  
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.