

Nietzer, Daniel; Henn, Karola; Islam, Chris-Gabriel; Krewer, Keno; Monzert, Carsten

Article

Hierarchisches Klassifizieren von Scannerdaten: ein Methodenvergleich mit Anwendung in der Verbraucherpreisstatistik

WISTA - Wirtschaft und Statistik

Provided in Cooperation with:

Statistisches Bundesamt (Destatis), Wiesbaden

Suggested Citation: Nietzer, Daniel; Henn, Karola; Islam, Chris-Gabriel; Krewer, Keno; Monzert, Carsten (2025) : Hierarchisches Klassifizieren von Scannerdaten: ein Methodenvergleich mit Anwendung in der Verbraucherpreisstatistik, WISTA - Wirtschaft und Statistik, ISSN 1619-2907, Statistisches Bundesamt (Destatis), Wiesbaden, Vol. 77, Iss. 1, pp. 67-81

This Version is available at:

<https://hdl.handle.net/10419/313330>

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.

HIERARCHISCHES KLASSIFIZIEREN VON SCANNERDATEN: EIN METHODEN- VERGLEICH MIT ANWENDUNG IN DER VERBRAUCHERPREISSTATISTIK

Daniel Nietzer, Karola Henn, Chris-Gabriel Islam, Keno Krewer,
Carsten Monzert

📌 **Schlüsselwörter:** Maschinelles Lernen – Hierarchisches Lernen –
Verbraucherpreisindex – Klassifizierung – COICOP

ZUSAMMENFASSUNG

Die Nutzung von Scannerdaten in der Verbraucherpreisstatistik hat viele Vorteile, bringt allerdings auch einige Herausforderungen mit sich. Eine der Herausforderungen ist die Klassifizierung der Artikel nach dem vom Verbraucherpreisindex verwendeten Klassifikationssystem COICOP. Aufgrund der großen Datenmenge bei Scannerdaten ist eine automatisierte Klassifizierung mittels maschinellen Lernens zwingend, um die Daten effizient in der amtlichen Statistik verwenden zu können. Dieser Beitrag analysiert einen Ansatz aus der aktuellen Forschung, der die zugrunde liegende hierarchische Struktur der COICOP-Klassifikation bei der automatisierten Klassifizierung berücksichtigt. Außerdem werden die Auswirkungen von Fehlklassifizierungen auf den Verbraucherpreisindex untersucht.

📌 **Keywords:** *machine learning – hierarchical learning – consumer price index – classification – COICOP*

ABSTRACT

The use of scanner data in consumer price statistics offers many benefits but also presents certain challenges. One such challenge is the classification of articles according to the COICOP system, which is used for the consumer price index (CPI). Due to the sheer volume of scanner data, automated classification using machine learning is essential to be able to use the data efficiently in official statistics. This article examines a current research approach that incorporates the underlying hierarchical structure of the COICOP classification during the automated classification of articles. The impact of misclassification on the CPI is also analysed.

Daniel Nietzer

hat Volkswirtschaftslehre an den Universitäten Göttingen und Mannheim studiert. Er schrieb seine diesem Beitrag zugrunde liegende Masterarbeit im Referat „Methoden und Kommunikation in der Preisstatistik“ des Statistischen Bundesamtes.

Karola Henn

ist Diplom-Finanz- und Wirtschaftsmathematikerin und Referentin im Referat „Methoden und Kommunikation in der Preisstatistik“ des Statistischen Bundesamtes. Hauptsächlich beschäftigt sie sich mit der Einführung von Scannerdaten in den Verbraucherpreisindex, ein Schwerpunkt ist dabei die Klassifizierung von Scannerdaten.

Dr. Chris-Gabriel Islam

hat Wirtschaftsmathematik und Lehramt in Mathematik und Geschichte in Darmstadt und Mexiko-Stadt studiert sowie in Göttingen und Hasselt (Belgien) promoviert. Er hat bis 2024 das Projekt „Scannerdaten in der Verbraucherpreisstatistik“ geleitet.

Dr. Keno Krewer

hat Physik in Mainz studiert und am Max-Planck-Institut für Polymerforschung promoviert. Er verantwortete die Methodik des Projekts „Scannerdaten in der Verbraucherpreisstatistik“ im Statistischen Bundesamt.

Carsten Monzert

hat Volkswirtschaftslehre in Frankfurt am Main und Mainz studiert. Er ist Data Engineer im Referat „Datenlabor“ des Statistischen Bundesamtes.

Danksagung

Die diesem Artikel zugrunde liegende Masterarbeit ist am Lehrstuhl für Ökonometrie der Georg-August-Universität Göttingen entstanden. In diesem Zusammenhang bedanken sich Autorin und Autoren herzlich bei Prof. Dr. Helmut Herwartz für die Kooperation.

1

Einleitung

Um den Verbraucherpreisindex (VPI) in Deutschland zu ermitteln, erheben vorwiegend die Statistischen Ämter der Länder dezentral die Preise vor Ort. Neue Datenquellen und technische Weiterentwicklungen ermöglichen es, in Digitalisierungsprojekten die Datenerhebung angesichts dynamischer und individualisierter Preissetzungen weiterzuentwickeln. So werden beispielsweise seit mehreren Jahren Preise automatisiert per Web Scraping im Internet erhoben (Blaudow/Ostermann, 2020). Auch Scannerdaten als digital vorliegende Datenquelle bieten sich an, um die Datenerhebung weiterzuentwickeln. Scannerdaten können die Preismessung zeitlich und räumlich sowie in Bezug auf den Sortimentsumfang ausweiten. Ziel ist es, die Genauigkeit und Qualität des Verbraucherpreisindex zu verbessern, indem die Preiserhebung erweitert wird und sie damit die Preisentwicklung im Einzelhandel repräsentativer darstellt (Bieg, 2019).

Bei Scannerdaten handelt es sich um Scannerkassendaten, die zu jedem verkauften Artikel Absatz, Umsatz, zugehörige Artikelinformationen und Regionalangaben für einen bestimmten Zeitraum enthalten. Um sie in der laufenden Statistikproduktion zu nutzen, fordert das Statistische Bundesamt Datenlieferungen von Einzelhandelsunternehmen an, die dann größtenteils von einer zentralen Stelle der Handelskette geliefert werden. Für die Anforderung dieser Transaktionsdaten wurde zunächst das Preisstatistikgesetz geändert. Dabei wurde insbesondere § 7b Absatz 3 Preisstatistikgesetz, der die Anforderungen von Transaktionsdaten regelt, aufgenommen. Außerdem wurde über § 3 Absatz 1a Preisstatistikgesetz ausdrücklich die Übermittlung von Umsatz- und Absatzinformationen berücksichtigt. Eine Besonderheit unter anderem bei den Scannerdaten ist auch die Regelung in § 7c Preisstatistikgesetz, die eine statistische Auswertung auch zu anderen Zwecken als der Preisstatistik ermöglicht. So werden Scannerdaten bereits für [experimentelle Konjunkturstatistiken](#) verwendet.

Scannerdaten werden in einer eigens entwickelten Fachanwendung aufbereitet, dabei werden Angaben plausibilisiert, imputiert und klassifiziert, um anschließend Preisindizes für die entsprechenden Güter zu berech-

nen. Da der Verbraucherpreisindex nach der COICOP¹ klassifiziert wird, müssen auch Scannerdaten nach der COICOP gegliedert ausgewertet werden. Dieser Schritt der Klassifizierung erfolgt aktuell in einem mehrstufigen Verfahren, das unter anderem Maschinelles Lernen beinhaltet. Auf diesen Schritt baut die in diesem Beitrag betrachtete Machbarkeitsstudie auf und versucht, das aktuelle Klassifizierungsverfahren durch Verfahren des hierarchischen Klassifizierens zu verbessern. Auch das hierarchische Klassifizieren hat zum Ziel, die Genauigkeit und Qualität des Verbraucherpreisindex zu verbessern, indem die Klassifizierung weiter verbessert wird.

Die Integration von Scannerdaten in den Verbraucherpreisindex führt dazu, dass deutlich mehr Artikel in den Index einbezogen werden können, da Scannerdaten das komplette Sortiment der verkauften Güter abbilden. Auch dehnen Scannerdatenlieferungen den möglichen Erfassungszeitraum erheblich aus und bilden damit die Preisänderungen umfassender und somit realistischer ab. Während die bisherigen Vor-Ort-Erhebungen die Preise in der Regel nur an einem Tag im Monat erfassen, ist bei Scannerdaten geplant, für die Indexberechnung die Daten der ersten beiden vollen Wochen eines Monats zu nutzen. Damit können beispielsweise Sonderangebote, die nur in einer Woche oder nur an bestimmten Wochentagen gelten, besser berücksichtigt werden.²

Um Scannerdaten im Verbraucherpreisindex produktiv nutzen zu können, müssen die Daten entsprechend dem Klassifikationssystem der Verbraucherpreisstatistik klassifiziert werden. Eine korrekte Klassifizierung ist notwendig, um den Verbraucherpreisindex möglichst verzerrungsfrei abzubilden. Das aktuelle Vorgehen bei den Scannerdaten sieht vor, dass für die Klassifizierung der Scannerdaten ein flaches maschinelles Lernverfahren angewendet wird, das die Hierarchie der COICOP nicht nutzt. In der im Folgenden beschriebenen Machbarkeitsstudie wurde versucht, die Hierarchie der Klassifikation

1 Die Klassifikation der Verwendungszwecke des Individualverbrauchs (Classification of Individual Consumption by Purpose – COICOP) ist eine Klassifikation der Statistikabteilung der Vereinten Nationen zur Klassifizierung und Analyse der Konsumausgaben der privaten Haushalte, privaten Organisationen ohne Erwerbszweck und des Staates nach dem Verwendungszweck, siehe ec.europa.eu

2 Auch bei den Scannerdaten kann es vorkommen, dass Sonderangebote, die nur für einen kurzen Zeitraum gelten, in die Berechnung für diesen Zeitraum einfließen. Der Effekt ist aber immer geringer als bei einer Vor-Ort-Erhebung, bei der im Zweifelsfall ein Sonderangebot, das nur einen Tag gilt, den kompletten Erfassungszeitraum repräsentiert. Des Weiteren wäre es denkbar, in Zukunft die ersten drei oder vier vollen Wochen in den Preisindex einfließen zu lassen.

zu nutzen, um bei maschinellen Klassifizierungsschritten bessere Ergebnisse zu erzielen. Kapitel 2 stellt die Theorie und Methodik eines hierarchischen Klassifizierungsansatzes sowie die Ergebnisse der Machbarkeitsstudie vor. Kapitel 3 untersucht Abweichungen, die sich bei Anwendung verschiedener Klassifizierungsansätze im Vergleich zu einem Referenzindex ergeben. Abschließend folgen ein Fazit sowie eine Empfehlung für mögliche weitere Entwicklungsarbeiten.

2

Hierarchisches Klassifizieren als neuer Ansatz

2.1 Einführung in Maschinelles Lernen und Klassifizieren

Die COICOP-Klassifikation ist ein hierarchisches Klassifikationssystem, das bis zur vierten Ebene (COICOP-5-Steller) europäisch harmonisiert ist. Für den nationalen Verbraucherpreisindex werden die einzelnen Positionen tiefer gegliedert und erhalten einen zehnstelligen Code (10-Steller). Um Preisindizes aus Scannerdaten zu erstellen, wird jedem gelieferten Artikel³ eine entsprechende Klasse auf dieser Ebene zugeordnet. Klassifizierung von Daten ist ein typischer Anwendungsfall von Maschinellem Lernen, der auch bei Scannerdaten effiziente Ergebnisse liefert.⁴

„Machine Learning oder maschinelles Lernen ist ein Teilbereich der künstlichen Intelligenz. Der Schwerpunkt liegt dabei auf dem Trainieren von Computern, um aus Daten und Erfahrungen zu lernen und sich stets zu verbessern – anstatt explizit dafür programmiert zu werden.“ (SAP Deutschland, 2025) Zu unterscheiden sind zwei Teilbereiche des maschinellen Lernens: das unüberwachte Lernen und das überwachte Lernen.

³ Ein Datensatz enthält alle Verkäufe eines Artikels einer Filiale in einer Kalenderwoche. Das heißt, Artikel sind mehrfach in Datenlieferungen vorhanden, da die Artikel in unterschiedlichen Filialen verkauft werden. Außerdem sind die Artikel auch in Datenlieferungen weiterer Wochen vorhanden. Jeder Artikel muss allerdings nur einmal klassifiziert werden.

⁴ Zu weiteren Aspekten von Maschinellem Lernen in der amtlichen Statistik siehe Dumpert (2024).

Beim **unüberwachten Lernen** versucht die Maschine, in den Daten selbstständig Muster und Zusammenhänge zu erkennen. Hierbei gibt es keine Zielvorgabe, es handelt sich also um einen explorativen Ansatz.

Klassifizierungsaufgaben hingegen sind ein Beispiel für **überwachtes Lernen**, bei dem ein Algorithmus auf Grundlage bereits manuell klassifizierter Daten (Trainingsdaten) Muster erkennt, um neue Artikel in die richtige Kategorie beziehungsweise Klasse einzusortieren.

Bei dieser Art des Lernens werden Kenntnisse aus einem bekannten Datensatz genutzt, um für einen anderen unbekanntem Datensatz Vorhersagen zu treffen. Die Maschine lernt Zusammenhänge und Muster der Daten und wendet diese anschließend auf den unbekanntem Datensatz an. Für das Trainieren eines solchen Algorithmus wird ein Datensatz von Klassifizierungsbeispielen benötigt, welcher neben den Variablen, die für das Klassifizieren genutzt werden, die korrekten Klassen enthält. Dieser Datensatz wird in zwei Teile unterteilt: einen Trainings- und einen Testdatensatz. Die Teilung wird vorgenommen um zu verhindern, dass der Klassifikator im Training den Datensatz auswendig lernt (overfitting), aber im Echtbetrieb bei der Klassifizierung von neuen Daten schlechtere Ergebnisse liefert. Deswegen werden Klassifikatoren im Bereich des überwachten Lernens mit Trainingsdaten trainiert und anschließend mithilfe der Testdaten evaluiert. Dieses trainierte Modell wird als Klassifikator bezeichnet. Beispiele für Methoden des überwachten Lernens sind neuronale Netze, Support Vector Machines und das in dieser Untersuchung angewendete Verfahren Random Forest (James und andere, 2013).

Das hierarchische Klassifizieren ist eine Form des überwachten Lernens und hat in den letzten Jahrzehnten an Bedeutung gewonnen. Beim hierarchischen Klassifizieren wird die vorhandene Hierarchie des Klassifikationssystems genutzt, das heißt der Klassifikator klassifiziert sukzessive auf jeder Ebene beginnend von der obersten.

Bisher sind die meisten der heutzutage verwendeten Klassifikatoren flache Klassifikatoren, die die hierarchische Struktur des Klassifikationssystems nicht berücksichtigen und direkt Vorhersagen auf der untersten Ebene treffen. Vorbereitend auf die Einführung in den Verbraucherpreisindex werden flache Klassifikatoren genutzt, hierbei werden die beobachteten Artikel direkt einem 10-Steller zugeordnet, während beim hierar-

chischen Klassifizieren die Artikel schrittweise immer feineren Klassen zugeordnet werden. Die Zuordnung auf der 10-Steller-Ebene erfolgt erst im letzten Schritt. In der COICOP-Klassifikation bedeutet dies, dass beim hierarchischen Klassifizieren ein Granny-Smith-Apfel nicht direkt dem 10-Steller 0116130200 „Äpfel“ zugeordnet wird, sondern zuerst dem COICOP-2-Steller 01 „Nahrungsmittel und alkoholfreie Getränke“, dann dem COICOP-3-Steller 011 „Nahrungsmittel“. Die Zuordnung wird über die 4-, 5- und 7-Steller fortgesetzt, bis der Granny-Smith-Apfel im letzten Schritt dem 10-Steller 0116130200 „Äpfel“ zugeordnet wird. [↘ Grafik 1](#)

Hierarchische Klassifikatoren besitzen keinen grundsätzlichen Vorteil gegenüber flachen Klassifikatoren. Welches Verfahren besser geeignet ist, hängt von dem genutzten Klassifikator und Klassifikationssystem ab und ist für jeden Anwendungsfall zu untersuchen. Grundsätzlich kann der Ansatz des hierarchischen Klassifizierens eine Klassifizierung in zwei Aspekten verbessern: Zum einen könnten mehr Artikel zur korrekten Klasse zugeordnet werden. Zum anderen führen selbst falsche Zuordnungen zu geringeren Fehlern auf oberen Ebenen,

falls zumindest eine Oberklasse⁵ korrekt zugeordnet ist. Bei der Preisstatistik entspricht die oberste Ebene dem Verbraucherpreisindex. In dem in Grafik 2 dargestellten Fall könnte zwar der Teilindex der untersten Klasse durch eine Fehlklassifizierung verzerrt sein, die Verzerrung des Teilindex der gemeinsamen Oberklasse könnte aber reduziert werden. Von einer Fehlklassifizierung eines Artikels sind die in [↘ Übersicht 1](#) genannten Indizes betroffen.

Übersicht 1

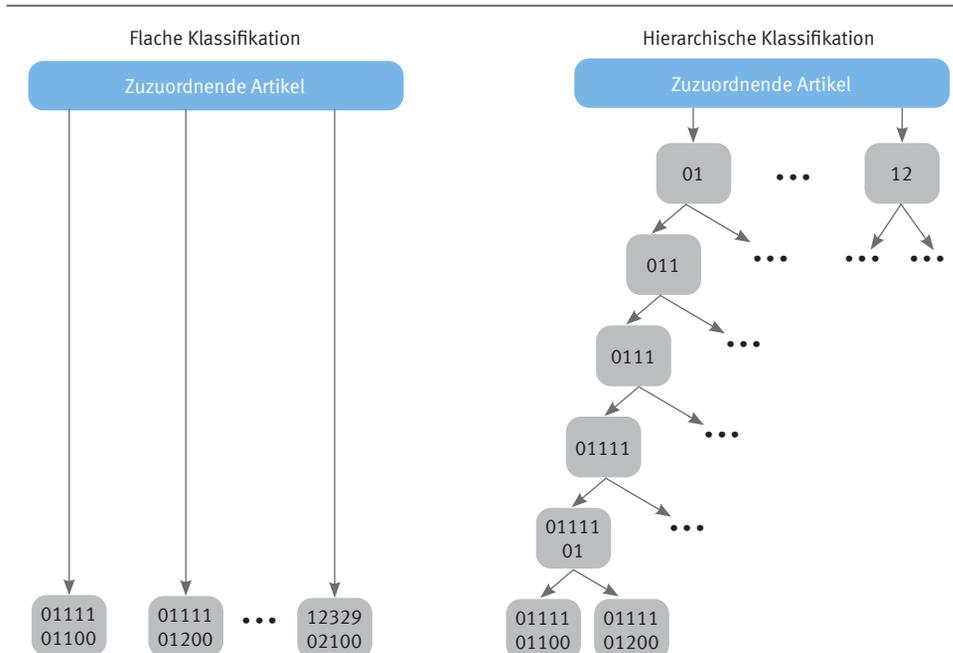
Relevante Indizes bei einer Fehlklassifizierung

Nummer	Bezeichnung	Erklärung
(1)	Spender-Index	Der Index der Klasse (10-Steller), in die der Artikel tatsächlich gehört
(2)	Empfänger-Index	Der Index der Klasse (10-Steller), in die der Artikel zugeordnet wurde
(3)	Spender-Aggregate	Indizes von Oberklassen, zu denen nur der Spender-Index beiträgt
(4)	Empfänger-Aggregate	Indizes von Oberklassen, zu denen nur der Empfänger-Index beiträgt
(5)	Gemeinsame Aggregate	Indizes von Oberklassen, die sowohl Spender als auch Empfänger enthalten

5 Im hierarchischen Klassifizieren bezeichnet der Begriff Oberklasse eine übergeordnete Klasse und der Begriff Unterklasse eine untergeordnete Klasse: Apfel ist eine Unterklasse der Oberklasse Früchte.

Grafik 1

Vergleichende Darstellung einer flachen und einer hierarchischen Klassifikation am Beispiel der COICOP¹



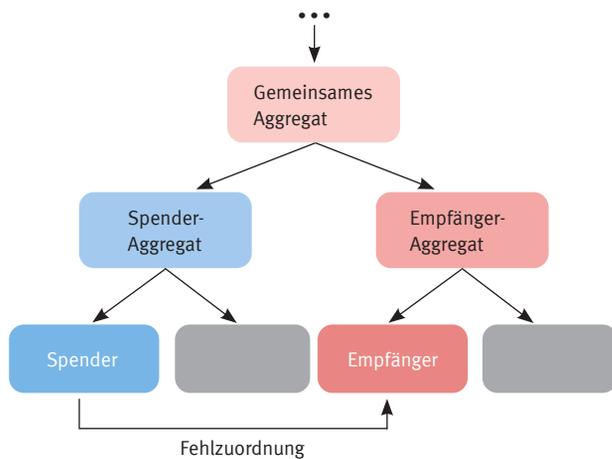
1 Classification of Individual Consumption by Purpose – Klassifikation der Verwendungszwecke des Individualverbrauchs.

Für die Indizes (1) und (2) der einzelnen Klassen gilt, dass nah verwandte Klassen häufig ähnliche Preisentwicklungen haben. Wenn sich die Preisreihen aus der Spender-Klasse ähnlich wie die der Empfänger-Klasse verhalten, ist die Auswirkung der Fehlzurordnung gering. Dies ist der erste Grund, warum Fehlzurordnungen mit geringen Abständen zwischen dem korrekten 10-Steller und dem zugeordneten 10-Steller nur wenig Einfluss auf die Qualität der Indizes haben.

Der zweite Grund liegt in der Aggregation: Je höher aggregiert wird, desto geringer sind die Auswirkungen der Fehlzurordnung einzelner Artikel, sofern die Zuordnung der Oberklassen korrekt erfolgt ist. Die Fehler im Spender- beziehungsweise im Empfänger-Aggregat erben ihr Verhalten von den Einzelindizes, im gemeinsamen Aggregat heben sich die Auswirkungen der Fehlzurordnung teilweise auf. [↘ Grafik 2](#)

Grafik 2

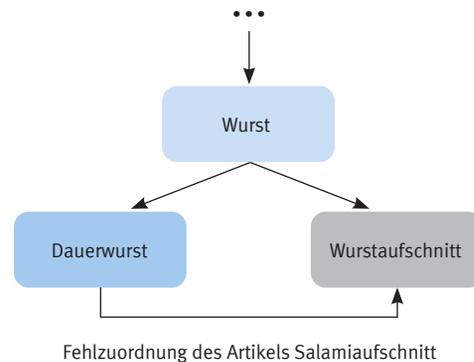
Auswirkung einer Fehlzurordnung eines Artikels aus dem Spender- zum Empfänger-10-Steller



[↘ Grafik 3](#) verdeutlicht dies mit einem Beispiel: Der Artikel „Salamiaufschnitt“ ist eigentlich eine Dauerwurst, wird aber fälschlicherweise dem Wurstaufschnitt zugeordnet. Eine Preisänderung des „Salamiaufschnitts“ wird jetzt dem Wurstaufschnitt zugerechnet, im Teilindex für Dauerwurst fehlt sie. Im gemeinsamen Aggregat „Wurst“ ist die Änderung wieder enthalten. Lediglich die Gewichtung aus dem neuen Teilindex, mit der die Preisänderung des Salamiaufschnitts eingeht, könnte verfälscht sein. Der Salamiaufschnitt geht fälschlicherweise mit dem Gewicht der Klasse Wurstaufschnitt multipliziert

Grafik 3

Beispielhafte Fehlzurordnung des Salamiaufschnitts



mit dem Anteil des Salamiaufschnitts an der Klasse des Wurstaufschnitts ein. Die Gewichte der Klassen sowohl in der Oberklasse als auch in der Unterklasse basieren innerhalb der Scannerdatenindizes auf Umsatzanteilen von Dauerwurst beziehungsweise Wurstaufschnitt. Ein Artikel mit gegebenem Umsatz wird also einen kleineren Umsatzanteil (kleines Gewicht) an einer umsatzstarken (schwergewichtigen) Klasse oder einen größeren Anteil (großes Gewicht) an einer umsatzschwachen (leichtgewichtigen) Klasse aufweisen. Daher ist das Gewicht des Artikels im gemeinsamen Aggregat auch bei Fehlzurordnung ähnlich. Im gemeinsamen Aggregat ist also der Effekt der Fehlzurordnung wieder teilweise korrigiert. Je näher die falsche Klasse mit der korrekten Klasse verwandt ist, desto früher wird ein gemeinsames Aggregat gebildet und werden die Auswirkungen der Fehlzurordnung reduziert. Daher führt eine nähere Verwandtschaft zwischen korrekten und falsch zugeordneten Klassen zu einer geringeren Verzerrung in den Indizes höherer Aggregate.

2.2 Typisierung hierarchischer Klassifikatoren

Bei hierarchischen Klassifikatoren werden mit lokalen und globalen Klassifikatoren zwei Ansätze unterschieden, hierarchische Strukturen zu berücksichtigen. Beim globalen Klassifikator (auch Big-Bang-Klassifikator genannt) wird ein einziger Klassifikator gebaut und trainiert, welcher die hierarchische Struktur der Daten berücksichtigt (Silla/Freitas, 2011). Bei lokalen Klassifikatoren werden auf unterschiedlichen Ebenen der Hierarchie Klassifikatoren trainiert, die jeweils den Output

aus den vorherigen Klassifikatoren übernehmen, um Unterklassen vorherzusagen. Es gibt drei verschiedene Arten von lokalen Klassifikatoren:

- › lokale Klassifikatoren je Klasse (LCN),
- › lokale Klassifikatoren je Oberklasse (LCPN) und
- › lokale Klassifikatoren je Ebene (LCL).

Bei einem LCN-Ansatz wird je Klasse ein binärer Klassifikator trainiert; in Teil (b) von [Grafik 4](#) sind jeweils die Einheiten markiert, für die eigene Klassifikatoren trainiert werden. Bei einem LCPN-Ansatz wird je Oberklasse ein Klassifikator gebildet, das ist in Teil (c) von Grafik 4 dargestellt. Dass für den LCL-Ansatz je Ebene ein Klassifikator trainiert wird, der zwischen allen Unterklassen entscheidet, zeigt Teil (d) von Grafik 4. Um den Gegensatz zu flachen Klassifikatoren zu verdeutlichen, ist dieser in Teil (a) von Grafik 4 abgebildet.

Globale und lokale Ansätze haben jeweils verschiedene Vor- und Nachteile. Die lokalen Klassifikatoren sind leicht umsetzbar, da sie auf bekannten flachen Klassifikatoren basieren. Ein Nachteil ist, dass Fehler vererbt werden können. Falls ein Artikel wie „Salami-aufschnitt“ auf der ersten Ebene der Hierarchie statt in den richtigen COICOP-2-Steller 01 „Nahrungsmittel und alkoholfreie Getränke“ in den COICOP-2-Steller 12 „Andere Waren und Dienstleistungen“ eingeordnet

wird, kann der Artikel nicht mehr dem korrekten 10-Steller zugeordnet werden.⁶ Für den globalen Klassifikator spricht, dass nur ein einziges Modell gebildet und trainiert wird. Dieses Modell ist allerdings relativ komplex und anwendungsspezifisch. Da die globalen Klassifikatoren so komplex sind, können traditionelle Evaluationsmetriken nicht genutzt werden, was einen Vergleich mit flachen Klassifikatoren erschwert. Im Weiteren liegt der Fokus auf den lokalen Klassifikatoren, insbesondere auf LCPN und LCN [Teile (c) und (b) der Grafik 4], die auch in der Literatur am häufigsten verwendet werden (Silla/Freitas, 2011).

2.3 Aktuelles Vorgehen

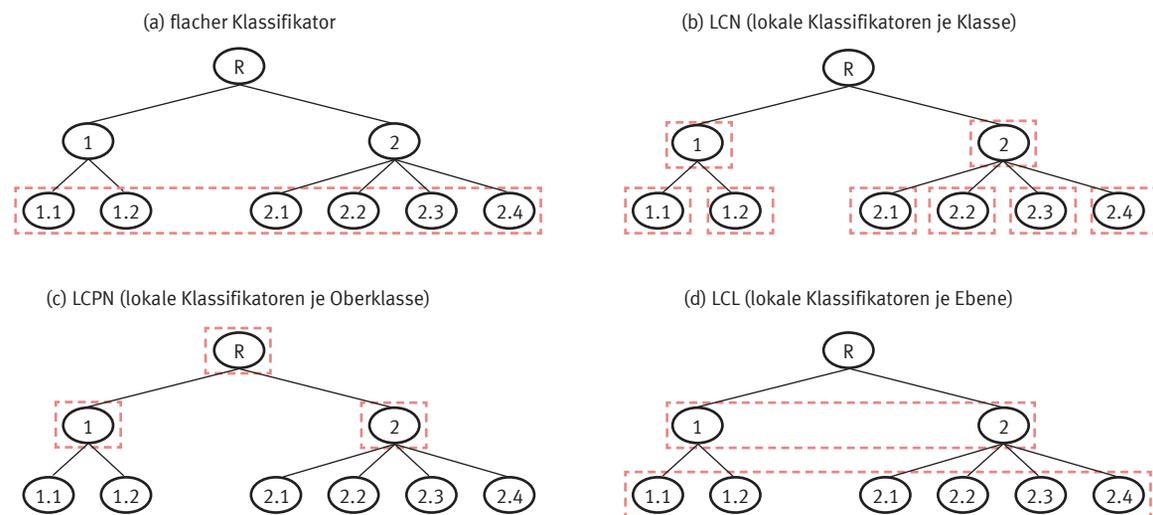
Das Projekt „Scannerdaten in der Verbraucherpreisstatistik“ bereitet die Einführung von Scannerdaten in den Verbraucherpreisindex zur nächsten Indexrevision vor. Hierfür ist geplant, die Scannerdaten der 23 umsatzstärksten Handelsketten⁷ aus dem Bereich Lebensmitteleinzelhandel und Drogeriemärkte zu nutzen. Um Verzerrungen bei der Berechnung von Scannerdatenindizes zu vermeiden, werden speziell für die Nutzung von Massendaten entwickelte Indexformeln verwendet. Um Scannerdaten in den Verbraucherpreisindex zu inte-

6 Grund hierfür ist, dass ein Top-down-Prädiktionsansatz verwendet wurde.

7 Es werden Handelsketten einbezogen, die kumuliert mindestens 80% des Umsatzes aller Handelsketten abdecken.

Grafik 4

Verschiedene Arten von Klassifikatoren



grieren, wird der Index auf Basis des GEKS-Törnqvist-Preisindex⁸ berechnet.

Um Indizes mittels Scannerdaten zu berechnen, werden nach Dateneingang alle Artikel in einem mehrstufigen Verfahren klassifiziert, wobei auch Maschinelles Lernen eingesetzt wird. In der ersten Stufe werden Lookup-Tabellen genutzt, in denen bereits bekannte Identifikatoren aus den Scannerdaten COICOP-Klassen zugeordnet sind. So kann eine Zuordnung entweder direkt über händlerinterne Warengruppen oder über den Barcode⁹ erfolgen. Die verbleibenden Artikel werden durch ein maschinelles Lernverfahren zugeordnet. Verwendet wird hierbei das Verfahren Random Forest, das zu den flachen Klassifikatoren gehört. Zur Qualitätssicherung werden Artikel, bei denen das Lernverfahren eine geringe Vertrauenswahrscheinlichkeit¹⁰ aufweist, manuell nachklassifiziert.

↘ **Tabelle 1** zeigt die Struktur des Trainingsdatensatzes, der für jede Handelskette erstellt wird, um die Klassifikatoren zu trainieren. Anschließend erfolgt ein Rebalancing des Trainingsdatensatzes, um die Güte des Klassifikators zu verbessern.¹¹ Die Spalte „Rebalanced“

in Tabelle 1 gibt an, ob der Datensatz auch nach dem Rebalancing im Trainingsdatensatz verbleibt. Beträgt der Wert FALSE, so bleibt der Wert nach dem Rebalancing erhalten, ansonsten wird er entfernt. Die Klassifikatoren werden auf Basis der Test- und Trainingsdaten trainiert und evaluiert.

2.4 Vergleich und Zwischenfazit

Um die Güte der angewandten flachen und hierarchischen Klassifikatoren vergleichen zu können, werden Evaluationsmetriken¹² verwendet. Eine höhere Güte bedeutet, dass ein Klassifikator die Artikel häufiger den korrekten Klassen zuordnet. Die Ergebnisse sind auf 0 bis 1 normiert, wobei ein höherer Wert einem besseren Ergebnis entspricht. Die Evaluation von hierarchischen Klassifikatoren birgt zusätzliche Herausforderungen, da die Klassifizierung auf verschiedenen Ebenen stattfindet und traditionelle Metriken beziehungsweise Gütemaße dies nur unzureichend messen können. Zwei Metriken, die sich für die Bewertung von hierarchischen Klassifikatoren eignen und auf die weitere Metriken aufbauen, sind „Präzision“ und „Recall“. **Präzision** ist definiert als das Verhältnis zwischen der Anzahl der korrekt zu einer Klasse zugeordneten Artikel und der Anzahl der zu dieser Klasse zugeordneten Artikel. **Recall** bezeichnet das Verhältnis zwischen der Anzahl der korrekt zu einer Klasse zugeordneten Artikel und der Gesamtzahl der tatsächlichen Artikel dieser Klasse. Um die Leistung auf der 10-Steller-Ebene zu messen, wird als traditionelle Metrik die Metrik „F1“ genutzt, welche Präzision und Recall kombiniert. Der F1-Wert kann entweder als mikro- oder makro-Version berechnet werden. In der makro-Version wird der Durchschnitt über alle F1-Werte der einzelnen Klassen gebildet. Der F1-Wert der Klassen ist das

- 8 Der GEKS-Törnqvist-Index ist ein multilateraler Preisindex, der vom Statistischen Bundesamt und anderen nationalen Statistikämtern wie dem Office for National Statistics (ONS) eingesetzt wird.
- 9 Mit Barcode ist die Global Trade Item Number (GTIN) gemeint, die jeden Artikel international identifiziert.
- 10 Die Vertrauenswahrscheinlichkeit gibt an, wie wahrscheinlich es ist, dass für eine Beobachtung eine richtige Vorhersage getroffen wird. Beim verwendeten Random-Forest-Verfahren erhält man für jeden Artikel und jede COICOP eine Wahrscheinlichkeit, ob die zugeordnete Klasse korrekt ist.
- 11 Wenn die Klassen ungleich verteilt sind, verschlechtert sich die Güte des Klassifikators, vor allem für kleine Klassen (Fernández und andere, 2018). Beim Rebalancing wird durch die Techniken des Over- und Undersamplings die Ungleichverteilung der Klassen ausgeglichen, indem aus Mehrheitsklassen Datensätze entfernt (Undersampling) und Minderheitsklassen mit aus Datensätzen anderer Handelsketten synthetisierten Daten angereichert werden (Oversampling).

- 12 Es ist wichtig, mehrere Metriken zu verwenden, da diese die Güte der Klassifikatoren nach verschiedenen Kriterien bewerten.

Tabelle 1

Aufbau der Trainingsdatensätze (fiktive Daten)

Artikel-ID	Artikelname	Händlerkategorie	COICOP-10-Steller	Mehrwertsteuersatz	Rebalanced
54957608	BASMATI REIS	Speisekammer_Verarbeitete_Lebensmittel_Reis	0111101100	7	TRUE
30529904	Zitrone Bio	Lebensmittel_Frucht_Frisch	0116117100	7	FALSE
635343270	Kaffee Shampoo	Drogerie_Dusche_Shampoo	1213211100	19	FALSE
49967686	Apfel Shampoo	Drogerie_Dusche_Shampoo	1213211100	19	TRUE

Die Spalte „Rebalanced“ gibt an, ob der Datensatz nach dem Rebalancing im Trainingsdatensatz verbleibt (TRUE) oder nicht (FALSE).

harmonische Mittel aus Präzision und Recall. Der mikro F1-Wert wird direkt ohne Berücksichtigung der Klassen berechnet. Für eine Bewertung der hierarchischen Klassifizierung ist die makro-Version besser geeignet, da Falschzuordnungen in allen Klassen gleich gewichtet werden und deshalb eine spezifische Falschzuordnung in einer großen Klasse nicht die schlechte Zuordnung in vielen kleinen Klassen überdecken kann. Als weitere Evaluationsmetrik wird zusätzlich der „hierarchische F1“ (Kiritchenko und andere, 2006), eine hierarchische Version des mikro F1 genutzt, um die Klassifizierungsgüte auf allen Ebenen der COICOP-Klassifikation zu messen.

↘ **Tabelle 2** vergleicht den aktuell genutzten flachen Klassifikator (Random Forest) und zwei Ansätze des hierarchischen Klassifizierens (LCN und LCPN)¹³ auf sämtlichen Ebenen der COICOP-Klassifikation. Datengrundlage sind Trainingsdatensätze, welche jeweils ungefähr 13 000 Produkte enthalten. Insgesamt zeigt sich bei den verwendeten Evaluationsmetriken kein einheitliches Bild, wobei die Unterschiede zwischen den Klassifikatoren meist nur sehr gering sind. Während der flache Klassifikator im Vergleich minimal besser die hierarchische Struktur abbildet, ist der hierarchische LCN-Klassifikator minimal besser, um die unterste Klasse (10-Steller) vorherzusagen.

Neben makro F1 und hierarchischer F1 führt Tabelle 2 auch die beiden Metriken Recall und Präzision auf. Um ein vollständiges Bild zu erhalten, wird die Metrik makro F1 zusätzlich auf verschiedenen Ebenen ausgewertet,

angefangen von der 2-Steller-Ebene bis zur 7-Steller-Ebene. Damit lässt sich feststellen, auf welcher Ebene sich die Klassifizierungen unterscheiden. Da der flache Klassifikator grundsätzlich nur auf der untersten Ebene die Klassen vorhersagen kann, wurde für die Auswertung angenommen, dass die Vorhersagen des flachen Klassifikators konsistent auf höheren Ebenen sind. Wenn also ein Artikel dem 10-Steller 0112710200 „Dauerwurst“ zugeordnet wird, erhält dieser den COICOP-2-Steller 01 „Nahrungsmittel und alkoholfreie Getränke“, den COICOP-3-Steller 011 „Nahrungsmittel“ und auf den folgenden Ebenen die entsprechenden Klassen.

Die F1-Werte für die verschiedenen Ebenen zeigen, dass der LCN-Ansatz für alle Ebenen den flachen Klassifikator und auch den LCPN-Klassifikator übertrifft, der LCPN-Klassifikator liefert vergleichbare Ergebnisse zum flachen Klassifikator. Besonders deutliche Differenzen zeigen sich auf der 5-Steller-Ebene, was insbesondere für die Vergleichbarkeit innerhalb der Europäischen Union interessant sein könnte. Der dort verwendete Harmonisierte Verbraucherpreisindex (HVPI) verwendet als tiefste Gliederung diese 5-Steller-Ebene.

Für die Klassifizierung im Rahmen des Verbraucherpreisindex ist die Metrik Präzision ausschlaggebend. Während sich bei konstanter Präzision durch einen niedrigen Recall die Varianz im Index erhöht, führt eine niedrige Präzision bei konstantem Recall zu einer systematischen Verzerrung im Index (Spackman und andere, 2023). Bezogen auf die Präzision übertreffen die hierarchischen Klassifikatoren den flachen Klassifikator. Für den Recall schneidet der LCN besser und der LCPN schlechter als der flache Klassifikator ab. Zwischen LCN- und LCPN-

13 Für den LCN und den LCPN wurde für jeden lokalen Klassifikator ein Random Forest genutzt.

Tabelle 2
Durchschnittliche Ergebnisse von Klassifikationsverfahren nach verschiedenen Evaluationsmetriken

Evaluationsmetrik \ Klassifikator	Flacher Klassifikator	LCN (lokale Klassifikatoren je Klasse)	LCPN (lokale Klassifikatoren je Oberklasse)
makro F1 (10-Steller)	0,861	0,864	0,860
hF (10-Steller)	0,951	0,950	0,946
Recall (10-Steller)	0,880	0,881	0,876
Präzision (10-Steller)	0,868	0,872	0,872
makro F1 (2-Steller)	0,938	0,942	0,937
makro F1 (3-Steller)	0,944	0,947	0,942
makro F1 (4-Steller)	0,944	0,947	0,940
makro F1 (5-Steller)	0,893	0,902	0,895
makro F1 (7-Steller)	0,869	0,874	0,869

Ansatz ist die Präzision gleich, der LCN-Ansatz erreicht jedoch einen höheren Recall. Der Unterschied zwischen den beiden hierarchischen Klassifikatoren beim F1-Wert wird deswegen vom Recall getrieben.

Zusammenfassend sind die Unterschiede zwischen dem flachen Klassifikator und den hierarchischen Klassifikatoren gering. Allerdings führt der LCN-Klassifikator zu leicht besseren Ergebnissen für fast alle genutzten Metriken, auch wenn die Verbesserung nur geringfügig ist.

Über die Machbarkeitsstudie hinaus besteht weiterer Optimierungsspielraum durch ein Rebalancing auf höheren Ebenen und eine individuelle Parameteroptimierung für jeden einzelnen Klassifikator innerhalb eines lokalen Ansatzes. Diese beiden Maßnahmen könnten die Güte der hierarchischen Klassifikatoren, aber auch – in Bezug auf das Rebalancing der Klassen – des flachen Klassifikators weiter verbessern.

3

Auswirkungen von Fehlklassifizierung auf die Scannerdatenindizes

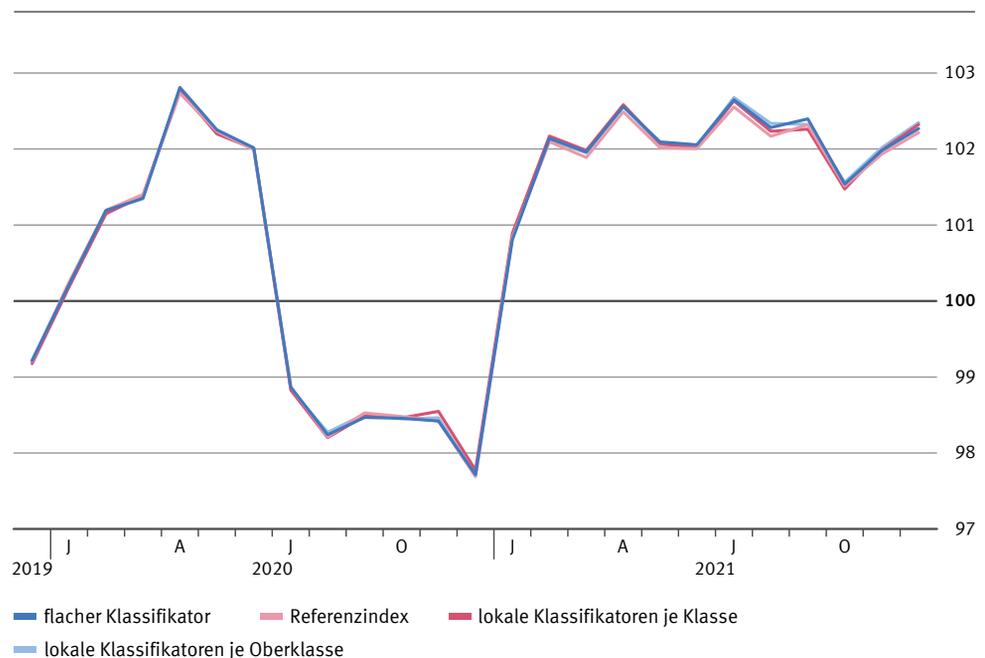
3.1 Vorgehen

Das Ziel der Klassifizierung von Scannerdaten ist es, den Verbraucherpreisindex möglichst verzerrungsfrei abzubilden, indem die Artikel in die korrekten 10-Steller der COICOP zugeordnet werden. Eine höhere Qualität in der Klassifizierung führt zu einer geringeren Verzerrung¹⁴. Im Folgenden wird aufgezeigt, wie sich die verschiedenen Klassifizierungsansätze auf einen Index basierend auf Scannerdaten auswirken. Dazu werden Scannerdatenindizes über verschiedene Arten der Klassifizierung erstellt, um die Auswirkung der Fehlklassifizierung zu isolieren. Würde man den Verbraucherpreisindex als Referenzindex nutzen, wäre es nicht möglich, den Einfluss der Fehlklassifizierung zu bestimmen, da der Verbraucherpreisindex auf anderen Daten aus der

14 Und mathematisch gesehen auch zu einer geringeren Varianz.

Grafik 5

Auswirkungen unterschiedlicher Klassifikatoren auf Scannerdatenindizes 2020 = 100



traditionellen Erhebung basiert.¹⁵ Daher werden die Indizes mit einem als unverfälscht angenommenen Scannerdatenindex verglichen, dessen Klassifizierung manuell erfolgt ist (Referenzindex). Der Beobachtungszeitraum umfasst 25 Monate, von Dezember 2019 bis einschließlich Dezember 2021.¹⁶ Es werden lediglich Scannerdaten aus dem Geschäftstyp 04 „Discounter und Drogeriemärkte“ genutzt. Diese enthalten wöchentliche Verkaufszahlen (Absatz) und den Umsatz jedes Artikels, der in einer Filiale verkauft wurde.

Für den Vergleich werden Indizes auf Basis der in Tabelle 2 genannten Klassifikatoren berechnet und mit dem Referenzindex verglichen. Der Referenzindex basiert, wie die anderen Indizes, ausschließlich auf Artikeln, die im manuell erstellten Trainingsdatensatz enthalten sind. Während für die zu vergleichenden Indizes jeweils die Zuordnung des jeweiligen Klassifikators

genutzt wird, wird für den Referenzindex der manuell zugeordnete Trainingsdatensatz¹⁷ verwendet und als korrekt angenommen. Dies gewährleistet, dass mögliche Abweichungen zwischen dem Referenzindex und den Indizes auf Basis der Klassifikatoren ausschließlich auf die Fehlklassifizierung zurückzuführen sind. In diesem Zuge werden Indizes auf 10-Steller-Ebene berechnet und durch ein gewichtetes arithmetisches Mittel auf höhere Ebenen aggregiert. Die Ergebnisse der verschiedenen Klassifikatoren zeigt [Grafik 5](#).

Um die Unterschiede zwischen den Scannerdatenindizes basierend auf den verschiedenen Klassifikatoren zu verdeutlichen, zeigt [Grafik 6](#) die absolute Abweichung jeweils zum Referenzindex für die verschiedenen Klassifikatoren.

Um das Ausmaß der Fehlklassifizierung auf den Scannerdatenindex zu bewerten, dienen die Grenzen, die das Statistische Amt der Europäischen Union (Eurostat) zur Definition eines erheblichen Methodenwechsels nutzt,¹⁸ als Orientierung. Laut der EU-Verordnung han-

15 Eine Abweichung findet nur statt, wenn der falsch zugeordnete Artikel eine andere Preisentwicklung hat als die Klasse, welcher er zugeordnet wurde.

16 Der GEKS-Törnqvist-Index benötigt einen längeren Zeitraum zur Berechnung der Indizes, gängig ist dabei ein 25-Monats-Fenster. In der vorliegenden Machbarkeitsstudie können für alle betrachteten Monate die gleichen Daten aus einem beliebigen Zeitfenster verwendet werden. Im Produktionsbetrieb können nur Daten der Vergangenheit verwendet werden, sodass jeweils ein 25-Monats-Fenster in der Vergangenheit verwendet wird.

17 Es wird nicht der gesamte Trainingsdatensatz, sondern nur ein Test-Split genutzt. Dies sind 25 % des gesamten Trainingsdatensatzes. Durch die geringe Anzahl der berücksichtigten Artikel ist der Index nicht mehr repräsentativ.

18 Artikel 2 Nummer 21 der Verordnung (EU) 2016/792.

Grafik 6

Absolute Abweichung der Scannerdatenindizes vom Referenzindex in Indexpunkten

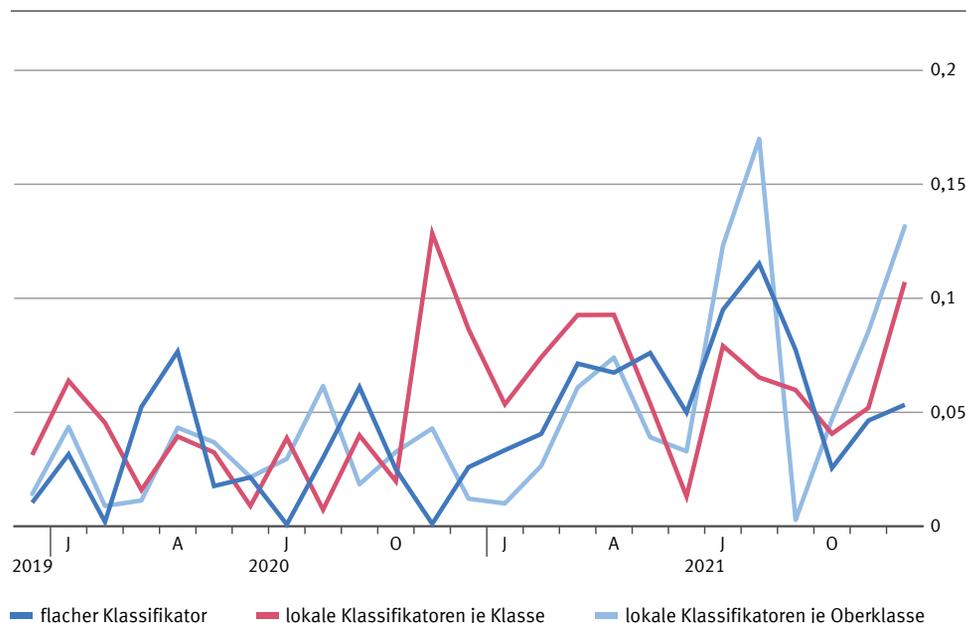


Tabelle 3

Angesetzte Grenzwerte für die Indexabweichung

	Verbraucherpreisindex	2-Steller	3-Steller	4-Steller	5-Steller	10-Steller
	Prozentpunkte					
Bund	0,1	0,3	0,4	0,5	0,6	1,0
Länder	0,2	0,6	0,8	1,0	1,2	2,0

Fett hervorgehobene Werte: bedeutender Methodenwechsel laut Verordnung (EU) 2016/792; übrige Werte in Anlehnung an die Verordnung.

delt es sich um einen bedeutenden Methodenwechsel, wenn die Veränderungsrate zum Vorjahresmonat eines Index oder eines Teilindex in einem beliebigen Zeitraum die fett hervorgehobenen Werte in [Tabelle 3](#) überschreitet. Die vorgegebenen Grenzen gelten bis zur 5-Steller-Ebene und sind auf nationale Ergebnisse ausgerichtet. Um auch für die 10-Steller und für die Bundesländerebene Grenzwerte zu erhalten, wird für den 10-Steller der Grenzwert 1,0 gesetzt, für die Bundesländerebene wird eine Verdopplung der Grenzwerte angesetzt. Die Grenzwerte auf Bundesländerebene und auf niedrigeren Ebenen der Güterklassifikation sind höher angesetzt, da weniger Artikel in den einzelnen Klassen enthalten sind: Dies kann zu einer höheren Varianz bei gleichbleibender Güte führen, da der Einfluss einzelner Artikel auf das Aggregat höher ist.

3.2 Ergebnisse

Wie Grafik 5 zeigt, sind die Unterschiede bei den Scannerdatenindizes, die auf Grundlage unterschiedlicher Klassifizierungsverfahren berechnet wurden, auf Bundesebene in der grafischen Darstellung kaum zu erkennen. Eine Auswertung der durchschnittlichen Abweichung vom Referenzindex unter Berücksichtigung der in Tabelle 3 aufgeführten Grenzwerte zeichnet jedoch ein differenzierteres Bild. Auf der 10-Steller-Ebene wird im betrachteten Zeitraum der Grenzwert für die Indizes auf Bundesebene auf Basis des flachen und des LCPN-Klassifikators bei 50 % und für den Index auf Basis des LCN-Klassifikators bei 44,87 % der Klassen in mindestens einem Monat überschritten. Alle drei Indizes überschreiten den Grenzwert auf der Scannerdaten-Gesamtindex-Ebene in mindestens einem Monat. Der Anteil der Klassen, die mindestens in einem Monat den Grenzwert überschreiten, ist bei den hierarchischen Klassifikatoren auf allen Ebenen niedriger oder gleich groß wie beim flachen Klassifikator, mit Ausnahme der 3-Steller-Ebene. [Tabelle 4](#)

Ähnliche Beobachtungen wie auf der Gesamtindexebene lassen sich auch bei den Scannerdatenindizes auf Bundesländerebene beobachten. Die Indizes auf

Tabelle 4

Scannerdatenindizes auf Basis unterschiedlicher Klassifikationsverfahren mit mindestens einem Monatsergebnis im Zeitraum Dezember 2019 bis Dezember 2021 außerhalb der Grenzwerte

Ebene	Klassen	Flacher Klassifikator		LCN (lokale Klassifikatoren je Klasse)		LCPN (lokale Klassifikatoren je Oberklasse)	
	Anzahl		%	Anzahl	%	Anzahl	%
Scannerdaten-Gesamtindex	1	1	100	1	100	1	100
2-Steller	7	3	42,68	2	28,57	2	28,57
3-Steller	10	2	20	3	30	3	30
4-Steller	24	8	33,33	6	25	6	25
5-Steller	81	46	56,79	43	53,09	45	55,56
10-Steller	234	117	50	105	44,87	117	50

Tabelle 5

Durchschnittliche Abweichung aufgrund von Fehlklassifizierung

Ebene	Insgesamt			Oberhalb des Grenzwertes			Unterhalb des Grenzwertes		
	Flach	LCN	LCPN	Flach	LCN	LCPN	Flach	LCN	LCPN
Scannerdaten-Gesamtindex	0,044	0,054	0,047	0,115	0,118	0,142	0,041	0,048	0,034
2-Steller	0,126	0,090	0,091	0,852	0,454	0,448	0,067	0,063	0,065
3-Steller	0,122	0,099	0,102	0,864	0,605	0,604	0,065	0,071	0,075
4-Steller	0,173	0,146	0,154	1,047	1,057	1,024	0,107	0,100	0,108
5-Steller	0,492	0,407	0,433	2,231	2,124	1,935	0,142	0,147	0,147
10-Steller	0,623	0,555	0,599	2,627	2,666	2,573	0,224	0,217	0,230

Erläuterungen: Flach = Flacher Klassifikator; LCN = lokale Klassifikatoren je Klasse; LCPN = lokale Klassifikatoren je Oberklasse

Basis der hierarchischen Klassifikatoren haben einen kleineren Anteil an Klassen mit mindestens einem Monat oberhalb des Grenzwertes auf fast allen Ebenen. Ausnahmen auf Länderebene bilden lediglich die Scannerdaten-Gesamtindex-Ebene, bei der der LCPN-Klassifikator einen gleich hohen Anteil an Überschreitungen hat wie der flache Klassifikator, sowie die 2-Steller-Ebene, bei der dasselbe für den LCN-Klassifikator gilt.¹⁹

Weiterhin ist ebenfalls die durchschnittliche absolute Abweichung in Prozentpunkten vom Referenzindex relevant. [Tabelle 5](#) verdeutlicht, dass der LCN-Klassifikator auf allen Ebenen außer der Scannerdaten-Gesamtindex-Ebene zur geringsten Abweichung führt. [Tabelle 5](#) zeigt außerdem, welche durchschnittlichen Abweichungen sich ergeben, wenn für die Ermittlung einerseits ausschließlich die Ebenen berücksichtigt werden, bei denen mindestens ein Monat eine Indexabweichung über dem Grenzwert aufweist (oberhalb Grenzwert), sowie andererseits Ebenen, bei denen die Indexabweichung in keinem Monat den Grenzwert überschreitet (unterhalb Grenzwert). Die Qualitätsunterschiede zwischen den Klassifikatoren sind minimal.

4

Fazit und Ausblick

Die Auswertungen der Tests zeigen kein einheitliches Bild, sodass der Qualitätsvorteil eines bestimmten Klassifikators nicht valide erkennbar ist. Zusammengefasst sind die aktuell verwendete Methode des flachen Klassifizierens und die dargestellte Methode des hierarchischen Klassifizierens als qualitativ gleichwertig einzustufen. Stellt man den aktuellen Klassifizierungsprozess um, wird daher keine bedeutsame Verbesserung bei der Klassifizierung erwartet. Dennoch liefert die vorliegende Untersuchung wichtige Erkenntnisse für Projekte der amtlichen Statistik, bei denen flache Klassifikatoren bei hierarchischen Klassifikationsstrukturen genutzt werden, zum Beispiel für die Klassifikation nach der „Systematik der Einnahmen und Ausgaben der privaten Haushalte“ bei der Einkommens- und Verbrauchsstichprobe.²⁰ Der vorliegende Methodenvergleich zeigt Ansatzpunkte, hierarchische Klassifizierungsmethoden zu erproben, um die Qualität der Klassifizierung zu verbessern. Ebenso sollten die hier untersuchten Ansatzpunkte auch in der Preisstatistik weiterverfolgt werden, insbesondere sobald die Scannerdaten weiterer Einzelhandelsketten aus dem Lebensmittelbereich eingebunden sind. Da die Ergebnisse stark von der Qualität der durch die Einzelhandelsketten gepflegten Artikelbeschreibungen und Warengruppen abhängen, könnte das hierarchische Klassifizieren bei Handelsketten aus anderen Bereichen zu besseren Ergebnissen führen.

19 Der Fokus des Artikels liegt in der Interpretation der Bundesergebnisse, die Ergebnisse auf Ebene der Bundesländer werden daher nicht im Detail aufgeführt.

20 Die bisher noch offene Fragestellung, ob es sinnvoll ist, hierarchische Klassifikationsverfahren zu nutzen, nennt auch Dumpert (2021).

Aktuell befindet sich das Projekt zur Integration von Scannerdaten in den Verbraucherpreisindex in der Entwicklungs- und Aufbauphase. Es ist geplant, Scannerdaten mit der nächsten Indexrevision des Verbraucherpreisindex in den Produktionsbetrieb zu überführen. Damit kann die Preismessung zeitlich und räumlich sowie in Bezug auf den Sortimentsumfang ausgeweitet und die Preisentwicklung besser abgebildet werden. In der Folge werden Vor-Ort-Erhebungen im Lebensmittel- und Drogeriebereich größtenteils ersetzt.

Währenddessen wird weiterhin versucht, den aktuellen Klassifizierungsprozess zu verbessern. Hierzu wird die Qualität der manuell erstellten Trainingsdaten durch die Konzeption und Bereitstellung weiterer Werkzeuge erhöht sowie das bisherige Klassifikationssystem überarbeitet²¹. Von einem verbesserten Klassifizierungsprozess bei den Scannerdaten profitiert letztlich nicht nur die Verbraucherpreisstatistik, sondern auch weitere amtliche Statistiken, in denen die Verwendung von Scannerdaten geprüft wird. Dazu zählen aktuell der Einzelhandel (Koch/Erdemsiz, 2020), die Haushaltserhebungen und die Kaufkraftparitäten. [u](#)

21 Dies umfasst die Umstellung von der COICOP 1998 auf die COICOP 2018 und Umsetzung auf die nationale 10-Steller-Ebene.

LITERATURVERZEICHNIS

- Blaudow, Christian/Ostermann, Holger. *Entwicklung eines generischen Programms für die Nutzung von Web Scraping in der Verbraucherpreisstatistik*. In: WISTA Wirtschaft und Statistik. Ausgabe 5/2020, Seite 103 ff.
- Bieg, Matthias. *Nutzung von Scannerdaten in der Preisstatistik – eine Untersuchung anhand von Marktforschungsdaten*. In: WISTA Wirtschaft und Statistik. Ausgabe 2/2019, Seite 25 ff.
- Dumpert, Florian. *Machine Learning in der amtlichen Statistik – Ergebnisse und Bewertung eines internationalen Projekts*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2021, Seite 53 ff.
- Dumpert, Florian. *Maschinelles Lernen im Statistischen Bundesamt*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2024, Seite 17 ff.
- Fernández, Alberto/García, Salvador/Galar, Mikel/Prati, Ronaldo C./Krawczyk, Bartosz/Herrera, Francisco. *Learning from Imbalanced Data Sets*. 2018. DOI: [10.1007/978-3-319-98074-4](https://doi.org/10.1007/978-3-319-98074-4)
- James, Gareth/Witten, Daniela/Hastie, Trevor/Tibshirani, Robert. *An Introduction to Statistical Learning – with Applications in R*. 2013. DOI: [10.1007/978-1-0716-1418-1](https://doi.org/10.1007/978-1-0716-1418-1)
- Kiritchenko, Svetlana/Matwin, Stan/Nock, Richard/Famili, A. Fazel. *Learning and Evaluation in the Presence of Class Hierarchies: Application to Text Categorization*. In: Lamontagne, Luc/Marchand, Mario. *Advances in Artificial Intelligence*. Canadian AI 2006. DOI: [10.1007/11766247_34](https://doi.org/10.1007/11766247_34)
- Koch, Julia/Erdemsiz, Baran. *Einsatz von Scannerdaten während der COVID-19-Pandemie*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2020, Seite 96 ff.
- SAP Deutschland. *Was ist Machine Learning? Definition von Machine Learning im Detail*. [Zugriff am 14. Januar 2025]. Verfügbar unter: www.sap.com
- Silla, Carlos N./Freitas, Alex A. *A survey of hierarchical classification across different application domains*. In: *Data Mining and Knowledge Discovery*. Ausgabe 22/2011, Seite 31 ff. DOI: [10.1007/s10618-010-0175-9](https://doi.org/10.1007/s10618-010-0175-9)
- Spackman, William/DeVilliers, Greg/Ritter, Christian/Goussev, Serge. *Identifying and mitigating misclassification: A case study of the Machine Learning lifecycle in price indices with web-scraped clothing data*. 2023. [Zugriff am 9. Januar 2025]. Verfügbar unter: unece.org

RECHTSGRUNDLAGEN

Gesetz über die Preisstatistik (Preisstatistikgesetz – PreisStatG) in der im Bundesgesetzblatt Teil III, Gliederungsnummer 720-9, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 3 des Gesetzes vom 17. Januar 2024 (BGBl. I Nr. 13) geändert worden ist.

Verordnung (EU) 2016/792 des Europäischen Parlaments und des Rates vom 11. Mai 2016 über harmonisierte Verbraucherpreisindizes und den Häuserpreisindex sowie zur Aufhebung der Verordnung (EG) Nr. 2494/95 des Rates (Amtsblatt der EU Nr. L 135, Seite 11).

Herausgeber
Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung
Dr. Daniel Vorgrimler
Redaktion: Ellen Römer

Ihr Kontakt zu uns
www.destatis.de/kontakt

Erscheinungsfolge
zweimonatlich, erschienen im Februar 2025
Ältere Ausgaben finden Sie unter www.destatis.de sowie in der [Statistischen Bibliothek](#).

Artikelnummer: 1010200-25001-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2025
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.