



Die Berechnungsmethode zur Ermittlung der Nichtanerkennungsquote nach der Registerzählung 2021

Statistik Austria
Direktion Bevölkerung
Bereich Demographie und Gesundheit

Kurzfassung

Im Rahmen der für den Stichtag 31.10.2021 durchgeführten Wohnsitzanalyse der Registerzählung, wurde das seit dem Finanzjahr 2014 eingesetzte Modell, welches den Anteil an „Karteileichen“ aus der Menge der Verdachtsfälle auf Basis eines statistischen Standardverfahrens schätzt, geprüft und adaptiert. Dieses aktualisierte Modell wird ab dem Finanzjahr 2024 für die Berechnung der sogenannten Nichtanerkennungsquote eingesetzt.

Das Modell weist all jenen Verdachtsfällen hohe Nichtanerkennungswahrscheinlichkeiten zu, welche in ihren demographischen Merkmalen jenen Datensätzen sehr stark ähneln, die sich im Rahmen der vergangenen Wohnsitzanalysen als „Karteileichen“ erwiesen haben. Mit den zusätzlichen zehn Jahren an Daten wurde das bisherige Modell weiterentwickelt und mit den neuesten Daten rekaliert.

Inhalt

1	Einleitung	3
2	Definitionen und Begriffe	4
3	Ausgangslage.....	5
4	Logistisches Regressionsmodell	6
5	Gliederung der Grundgesamtheit nach Gemeindeklassen	6
6	Ermittlung der Schwellwerte zur Nichtanerkennung von Klärfällen	8
7	Validierung des Regressionsmodells	8
7.1	Verfügbare / verwendete Variable.....	9
7.2	Untersuchte Modelle	10
7.2.1	Logistische Regression	10
7.2.2	Gruppierung der Gemeinden.....	10
7.2.3	Gradient-Boosting-Verfahren	11
8	Zusammenfassung.....	12
9	Literaturverzeichnis	12

1 Einleitung

Im Zuge der Probezählung 2006 wurde erstmals das Konzept der Lebenszeichenanalyse zur Ermittlung der tatsächlichen Bevölkerungszahl entwickelt [1] und bei der jährlichen Bevölkerungsfeststellung für den Finanzausgleich sowie der Registerzählung 2011 und 2021 erfolgreich angewandt.

Dabei wird ausgehend vom Zentralen Melderegister versucht, jede Person mit einer gültigen Hauptwohnsitzmeldung durch andere Register zu bestätigen. Diese als sogenannte Lebenszeichen titulierten Einträge in Registern, wie beispielsweise jenem des Dachverbands der Sozialversicherungsträger, des Arbeitsmarktservice oder der Dienstgeberdaten des Bundes und der Länder, erheben in Verbindung mit dem Zentralen Melderegister einen derart starken Bestätigungsanspruch, dass diese Datensätze unmittelbar als Wohnbevölkerung der betreffenden Gemeinde gezählt werden, sofern sie nicht unter die gesetzliche 90- oder 180-Tage-Regel fallen ("Technische Nichtanerkennungen")¹.

Die Komplementärmenge zur oben beschriebenen Hauptgruppe an Datensätzen ("Zählfälle") stellt zwar einen kleinen, jedoch sehr untersuchungswürdigen Personenkreis dar: Es handelt sich demnach um Hauptwohnsitzmeldungen zu welchen im Rahmen des registerbasierten Erhebungsverfahrens keinerlei bestätigende Einträge in anderen Registern gefunden werden konnten. Diese im Sprachgebrauch der Registerzählung als "Klärfälle" bezeichneten Datensätze stellen zu einem gewichtigen – aber eben nicht zum vollständigen – Teil Löschfälle dar, welche den betreffenden Gemeinden als Wohnbevölkerung abzuerkennen sind.

Im Zuge der Probezählung 2006, der Registerzählung 2011 und 2021 wurde von Statistik Austria an diese Personenmenge per RSb-Brief² eine Benachrichtigung versandt, in welcher um eine Bestätigung des fraglichen Hauptwohnsitzes ersucht wurde. Hierdurch konnten die Löschfälle unmittelbar festgestellt werden.³

¹ Dabei handelt es sich um Bestimmungen des Registerzählungsgesetzes [2]: 90-Tage-Regel (§ 7 Abs. 3): Personen werden zum Stichtag gezählt, wenn sie sich um den Stichtag herum mehr als 90 Tage in Österreich aufhalten. 180-Tage-Regel (§ 7 Abs. 2): Personen, die um den Stichtag herum weniger als 180 Tage in einer Gemeinde gewohnt haben und danach wieder in die Gemeinde zurückkehren, aus der sie vorher gekommen sind, werden nicht in der Stichtagsgemeinde gezählt, sondern in der Gemeinde des früheren und auch späteren Hauptwohnsitzes.

² Ein RSb-Brief ("Rückscheinbrief") ist ein behördliches Schriftstück, das einer Bestätigung durch eine Empfänger:in bedarf.

³ Aufgrund des Registerzählungsgesetzes musste den betroffenen Gemeinden auch die Möglichkeit gegeben werden, für die so ausgelisteten zweifelhaften Hauptwohnsitze noch andere Lebenszeichen zu erbringen (schriftliche Bestätigungen der Bürger:innen). Der Vorgang kann folglich als recht erschöpfend bezeichnet werden.

Für die außerhalb einer Registerzählung stattfindenden Bevölkerungsfeststellungen ergibt sich damit die Ausgangslage, dass zwar die Menge an fraglichen Klärfällen alljährlich ermittelt werden kann, aus diesen jedoch die Menge an Löschfällen möglichst exakt – allerdings ohne eine tatsächliche Briefbefragung – abzuleiten ist. Wie bereits bei Festlegung der Formel zu Nichtanerkennungen von Hauptwohnsitzen nach der Registerzählung 2011, wird die Wahrscheinlichkeit der Löschung durch statistische Klassifikation (logistische Regression) geschätzt. Daraus wird dann die gemäß Finanzausgleichsgesetz zu ermittelnde Bevölkerungszahl berechnet. Da nun mit der Registerzählung 2021 erstmalig seit zehn Jahren neue Daten bezüglich den tatsächlichen Zähl- und Löschfällen zur Verfügung stehen, wurde das bisherige Modell überprüft und adaptiert. Die grundsätzliche Struktur des Modells wurde nicht verändert, lediglich Variablen und Parameter wurden angepasst.

2 Definitionen und Begriffe

Technische Nichtanerkennungen (TNA): Personen im ZMR, für die Verstorbenermeldungen vor dem Stichtag vorliegen bzw. die unter die Regelungen nach § 7 Abs. 2 oder 3 Registerzählungsgesetz fallen, werden nicht anerkannt. Dies sind z.B. 90-Tage-Fälle, Kit-Fälle, 180-Tage-Fälle¹.

Zählfall (ZF): Personen mit einer gültigen Hauptwohnsitzmeldung, die durch andere Register bestätigt sind

Klärfall (KF): Personen mit einer gültigen Hauptwohnsitzmeldung, die durch andere Register NICHT bestätigt sind.

Löschfall (LF): Personen aus der Menge der Klärfälle, für die keine Bestätigung des Hauptwohnsitzes vorliegt. Diese werden auch als „Karteileichen“ bezeichnet.

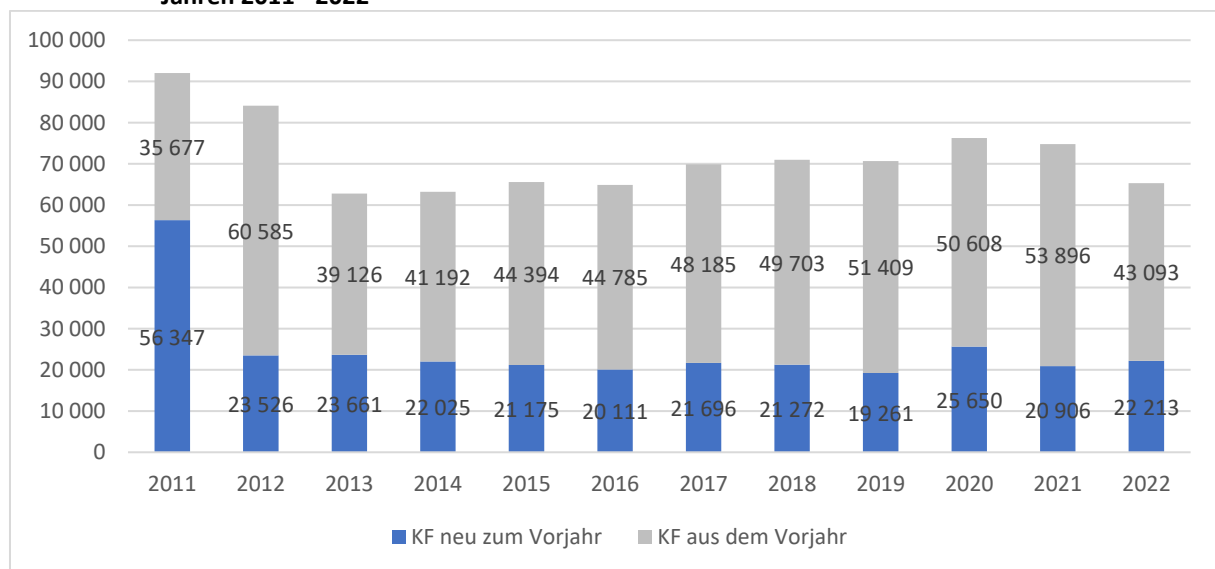
Anerkennungen (A): Personen aus der Menge der Klärfälle, für die letztendlich eine Bestätigung des Hauptwohnsitzes vorliegt bzw. die nach Anwendung des statistischen Modells anerkannt werden.

Nichtanerkennungen (NA): Personen aus der Menge der Klärfälle, für die keine Bestätigung des Hauptwohnsitzes vorliegt bzw. die nach Anwendung des statistischen Modells nicht anerkannt werden.

3 Ausgangslage

Technisch gesehen steht damit alljährlich eine Menge an Datensätzen mit exakten Adressdaten (und damit verbunden, in aller Regel einer recht genauen Erfassung der Wohnsituation) sowie einer Reihe von demographischen Personenmerkmalen zur Verfügung. Letzteres allerdings mit der Einschränkung der Datenlage, welche das Zentrale Melderegister in Bezug auf die Demographie zu bieten hat, da es sich dabei exakt um jene Meldesätze handelt, zu denen das registerbasierte Verfahren keine weiteren Informationen anreichern konnte.

Abbildung 1: Entwicklung der Klärfälle sowie des Anteils an im Vergleich zum Vorjahr neuen Meldungen in den Jahren 2011 - 2022



Q: STATISTIK AUSTRIA

Allein aus diesem jährlich erhaltenen Bestand lassen sich bereits bei bloßer zeitlicher Betrachtung, gewisse Schlüsse ziehen: prüft man, welche Datensätze bereits im Vorjahr ohne weitere Lebenszeichen (und damit ebenfalls Klärfälle) waren, so lässt sich unmittelbar feststellen, dass ein recht hoher Stock an unveränderten, alljährlich wiederkehrenden Verdachtsfällen in dieser Masse aufscheint. Die Information seit wann und wie oft eine Person nun Klärfall war steht nun in ausreichender Länge zur Verfügung und wurde in das neue Modell mit aufgenommen, was aufgrund der kurzen Zeithistorie für das bisherige Modell nicht möglich war.

Ergebnisse eines Population Stability Index (PSI) ergaben eine signifikante Veränderung der Verteilung für einige der verwendeten Inputvariablen des bisherigen Modells. Aufgrund dieser strukturellen Veränderung der Basisdaten im Laufe der Zeit und aufgrund mangelnder

historischer Information, hat man entschlossen, das Jahr 2006 aus der Datenbasis für die Neuentwicklung zu entfernen und nur die Jahre 2011 und 2021 heranzuziehen.

Mit Hilfe dieses Datenbestandes von etwa 167 000 Einzeldatensätzen lassen sich nun Rückschlüsse auf die verborgene Struktur der Löschfälle ziehen, indem alle enthaltenen Merkmale auf ihre Eigenschaft hin untersucht werden, ob sie Anerkennungen von Nichtanerkennungen zu trennen vermögen.

4 Logistisches Regressionsmodell

Die Entscheidung für ein Regressionsmodell fußte auf einer Reihe von Gründen. Zum einen eignete sich die vorliegende historische Datenbasis in ihrer Ausprägungsstruktur gut für einen derartigen Ansatz.

Nachvollziehbarkeit und Transparenz stellen im Rahmen der Registerzählung bei allen Überlegungen stets wichtige Leitgedanken dar: Das Verfahren eines logistischen Modells stellt ein Standardwerkzeug aus dem statistischen Methodeninventar dar. Das umgesetzte logistische Modell hat die Form

$$\text{logit}(\pi) \equiv \log \left(\frac{\pi}{1-\pi} \right) = \alpha + \boldsymbol{\beta}' \mathbf{x}$$

wobei α die Schätzkonstante und der Vektor β der Form

$$\boldsymbol{\beta} = (\beta_1, \dots, \beta_s)'$$

die gesuchten Koeffizienten für die potentiellen Trennvariablen darstellen.

Erhält man nun in den Folgejahren einen Datensatz mit allen benötigten Trennvariablen, wie dieser im Zuge der registerbasierten Bevölkerungsfeststellung für den Finanzausgleich alljährlich vorliegt, so lassen sich die entsprechenden Wahrscheinlichkeiten auf Einzeldatensatzebene zuordnen.

5 Gliederung der Grundgesamtheit nach Gemeindeklassen

Obleich der Kern der Ermittlung der Nichtanerkennungsquote der beschriebene Ansatz eines Regressionsmodells ist, finden noch einige verfahrensinhaltliche Modifikationen statt, welche den Datenrealitäten in der Registerzählung Rechnung tragen.

Statt lediglich ein einziges Regressionsmodell für Gesamtösterreich abzuleiten, wurden die Gemeinden in unterschiedliche Gruppen eingeteilt, sodass das Verfahren dem unterschiedlichen Naturell der einzelnen Gemeinden explizit Rechnung trägt. Die Gruppierung erfolgte für kleine und mittlere Städte und Gemeinden auf Basis der Anzahl der Klärfälle pro Gemeinde, Wien wurde aufgrund struktureller Unterschiede als gesonderte Gruppe kategorisiert. Somit sind die Modelle abgesichert gegen Änderungen von Gemeindegrenzen, die Zugehörigkeit zu einer Gruppe hängt für kleinere und mittlere Gemeinden/Städte lediglich von der Anzahl ihrer Klärfälle ab.

Nach einigen Versuchen mit anderen Vorgehensarten (nach Urbanisierungsgrad, geographischen Gegebenheiten etc.) erwies sich dieser einfache Ansatz als der zielführendste, die tatsächliche Realität abzubilden. Die historische Datenbasis wurde demnach in der Form aufgeteilt, dass sie fünf Gemeindegruppen widerspiegelt und schließlich dermaßen getrennt zur separaten Modellbildung übergeben.

Tabelle 1: Die Einteilung der Gemeinden nach der Anzahl ihrer Klärfälle in der historischen Datenbasis zur Behandlung mit unterschiedlichen Regressionsmodellen

Historische Datenbasis 2011 und 2021				
Gemeinden mit	Gruppe	Anzahl Gemeinden	Anzahl Klärfälle	Anzahl Löschfälle
10 000 KF und mehr	1	1	72 097	52 956
999 bis 9 999 KF	2	4	26 953	22 091
100 bis 999 KF	3	32	13 344	8 403
15 bis 99 KF	4	435	32 586	19 073
0 bis 14 KF	5	1 590	21 846	12 678
Gesamt		2 062	166 826	115 201

Q: STATISTIK AUSTRIA

Wie bisher wurden Fälle die in der letzten Briefbefragung (2021) geklärt wurden (anerkannt oder nicht anerkannt), werden in den Folgejahren analog behandelt. Zur Bewertung mit der geschätzten Löschwahrscheinlichkeit verbleiben also nur vollständig neue Klärfälle⁴. Diese Modifikation bringt neben ihrer sachlichen Notwendigkeit noch andere Vorteile wie eine

⁴ Es ist möglich, dass sich ein bestätigter Klärfall im Laufe der Zeit zur „echten Karteileiche“ entwickeln kann und somit die Person aufgrund dieser Methode des Fortschreibens der Anerkennungen weiter gezählt wird, dies wird jedoch in Kauf genommen und spätestens zum nächsten Registerzählungstermin bereinigt.

größere Modellstabilität in Bezug auf die historische Variabilität der gemeindespezifischen Löschoroten.

6 Ermittlung der Schwellwerte zur Nichtanerkennung von Klärfällen

Nachdem nun jedem Klärfall eine Löschorwahrscheinlichkeit zugeordnet wurde, stellt sich die Frage, welche Datensätze tatsächlich als Löschorfälle zu klassifizieren sind. Hierfür wird wie gefolgt vorgegangen.

Nach der Anwendung der von der historischen Datenbasis abgeleiteten Modelle auf der selbigen, werden die Datensätze pro Gemeindegruppe nach den erhaltenen Löschorwahrscheinlichkeiten absteigend sortiert, sowie die Anzahl der tatsächlich beobachteten Nichtanerkennungen N ausgezählt. Die Schwelle ab der künftig ein Löschorfall angenommen wird, stellt die dem an N -ter Stelle stehenden Datensatz zugewiesene Wahrscheinlichkeit dar.

Als Alternative zu den individuellen Schwellwerten der Löschorwahrscheinlichkeiten pro Gemeindegruppe wurde ein allgemeiner Wert für alle Modelle untersucht. Die vorgeschlagenen Werte stellten sich aus dem Mittelwert der neu kalkulierten Löschorwahrscheinlichkeiten, dem Schwellwert der gesamten Klärfälle, die man absteigend sortiert hat und einem Experten Wert von 70 % zusammen. Allerdings schwanken die Ergebnisse der Modellentscheidung und der Finalen Entscheidung extrem auseinander, sodass man aus Stabilitätsgründen bei den individuellen Schwellwerten verblieben ist.

7 Validierung des Regressionsmodells

Ziel des Regressionsmodells ist es, Löschorfälle bzw. „Karteileichen“ in einer Masse von Klärfällen zu identifizieren. Neben diesem „Hauptziel“ gibt es allerdings für die Modellbewertung nach klassischen statistischen Gesichtspunkten, den statistischen Kennzahlen und Tests, auch ein „Nebenziel“, das die Modellauswahl beeinflusst. Es sollen zwar möglichst die richtigen Fälle markiert werden, es ist aber auch wichtig (aus Sicht der betroffenen Gemeinden) die richtige Gesamtzahl der Löschorfälle je Gemeinde zu finden.

7.1 Verfügbare / verwendete Variable

Wie bereits in der Dokumentation zur Ermittlung der Nichtanerkennungsquote nach der Registerzählung 2011 [3] ausgeführt, ist bei der Variablenauswahl insbesondere darauf zu achten, dass zum Zeitpunkt der Durchführung der Analyse diese Informationen auch für den Bestand der Klärfälle zur Verfügung stehen. Da ein Klärfall per Definition ausschließlich im Melderegister und sonst in keinem anderen Register vorkommt, ist hier also für den jeweiligen Stichtag eine Einschränkung auf die Merkmale des Melderegisters gegeben. Außerdem gibt es Merkmale, die nicht die Person betreffen, sondern z.B. auf Ebene der Gemeinden aggregiert sind. Diese können auch aus anderen Quellen stammen und sind für alle Fälle einer Gemeinde gleich.

Zusätzlich zu den Variablen die bei der ursprünglichen Modellentwicklung getestet wurden, kamen nun auch Merkmale bezüglich der Klärfallvergangenheit der letzten Jahre hinzu. Diese wurden für die Variablenauswahl der einzelnen Modelle auf ihre Signifikanz untersucht.:

Tabelle 2: Variablenauswahl der potentiellen Trennvariablen

Variable	Erläuterung
ALTER	Alter zum Stichtag
ALTER_5	Alter in 5-Jahresgruppen
ALTER_GR	3 Altersgruppen (Kind, Erwerbsfähig, Pensionist)
ALTER_SQR	Alter zum Quadrat
ALTER_ZUZUG	Alter zum Zeitpunkt des Zuzugs nach Österreich
ANT_KF_GKZ	Anteil der Klärfälle an der GKZ
ANT_UEBER60	Anteil der über 60-jährigen in der Gemeinde
ANT_UNTER15	Anteil der unter 15-jährigen in der Gemeinde
ANZ_KF_JAHRE	Anzahl der Jahre in denen die Person ein KF war
ANZ_MELDAEND	Anzahl Meldeänderungen der letzten fünf Jahre = Übersiedlungen
ANZ_NTZ_OBJ	Anzahl NTZ = Wohnungen in einem Objekt = Gebäude
ANZ_WS	Anzahl Wohnsitze (inkl. Nebenwohnsitze)
ANZAHL_BEZIEHUNGEN	Nachweis/Anzahl von weiteren Beziehungen in der Beziehungstabelle
AZ_KF_GEM	Anzahl Klärfälle in der GKZ
BDL	Bundesland
DEM_GESCHL	Geschlecht
FAMST	Familienstand
GEBDAT	Geburtsdatum
GEBDAT_YEAR	Geburtsjahr
GEBSTAAT	Geburtsstaat
GEBSTAAT_GR	Aggregierte Kategorie Geburtsstaat: Österreich, Deutschland, EU13, EU13, EU-Rest, Kontinente
GEBSTAAT_GR2	Aggregierte Kategorie Geburtsstaat: Österreich, Deutschland, EU, Rest der Welt

GEBSTAAT_GR3	Aggregierte Kategorie Geburtsstaat: Österreich und Deutschland, Rest der Welt
KF_SEIT_JAHR	Jahr seitdem die Person ununterbrochen ein Klärfall ist
MAX_KF_JAHR	letztes Vorkommen als Klärfall
MIN_KF_JAHR	erstes Vorkommen als Klärfall
NUTS3	EU Klassifizierung - kleinere Regionen, teils schon Großstädte
PARTNER	Nachweis eines Partners in der Beziehungstabelle
PERS_M2_NTZ	Anzahl der Personen pro m ² bezogen auf Wohnung excl. NWS
PERS_M2_NTZ_NWS	Anzahl der Personen pro m ² bezogen auf Wohnung inkl. NWS
PERS_M2_OBJ	Kategorie Objekt: Personen pro m ² im Gebäude
STAATB	Staatsbürgerschaft
STAAT_GR	Aggregierte Kategorie Staatsbürgerschaft Österreich, Deutschland, EU13, EU13, EU-Rest, Kontinente

Q: STATISTIK AUSTRIA

7.2 Untersuchte Modelle

Um das bestmögliche Modell für die Ermittlung der Löschfälle zu finden und auch um eine Abschätzung für die Stabilität des Modells zu bekommen, wurden mehrere statistische Berechnungsmethoden untersucht. Zusätzlich zu den Modellen, die bei der initialen Entwicklung des Modells untersucht wurden (Boosting und Entscheidungsbäume) wurden als Alternative zur logistischen Regression ein Gradient-Boosting-Verfahren, sowie unterschiedliche Gruppierungsmethoden der Gemeinden untersucht.

Aufgrund der resultierenden Ergebnisse wurde entschieden weiterhin eine logistische Regression zu verwenden, wobei die unterschiedlichen Verfahren zu sehr ähnlichen Ergebnissen geführt haben.

7.2.1 Logistische Regression

In Punkt 4 sowie der Dokumentation zur Ermittlung der Nichtanerkennungsquote nach der Registerzählung 2011 [3] wurde bereits formal die logistische Regression behandelt.

Für die 5 Gemeinde- bzw. Klärfallgruppen wurden wieder in einem schrittweisen Verfahren für eine 70 %-ige Trainingsstichprobe jene Variablen entwickelt, die zur Erklärung beitragen. Das jeweilige Modell wurde anschließend anhand der 30 %-igen Validierungsstichprobe geprüft. Für alle fünf Gruppen konnten akzeptable Modelle ermittelt werden.

7.2.2 Gruppierung der Gemeinden

Zusätzlich zu der erweiterten Variablenauswahl wurden alternative Einteilungen in Gemeinde- bzw. Klärfallgruppen betrachtet. Allerdings entschloss man sich schlussendlich bei den ursprünglichen 5 Clustern zu bleiben.

Versucht wurde zunächst die Gemeinden mittels k-Means-Algorithmus zu gruppieren. Die entstandenen Cluster lieferten insgesamt keine Verbesserung der Trefferquote im Vergleich zu der aktuell verwendeten Einteilung.

Als nächstes wurden die Gemeinden anhand ihres Klärfallanteils in 5 Gruppen unterteilt. Wobei Wien aufgrund der hohen Anzahl an Klärfällen eine eigene Gruppe geblieben ist. Die restlichen Gemeinden wurden je nach KF-Anteil in 4 Gruppen eingeteilt. Die Trefferquote dieser Variante war allerdings noch schlechter als für das oben erwähnte k-Means-Clustering und wurde deshalb wieder verworfen.

Schließlich wurde untersucht ob die Grenzen der bisherigen KF-Gruppierung etwas angehoben werden soll. Da die Performance der Modelle für die Gemeindegruppen mit mehr KF bisher auch recht gut war, sollte dadurch mehr Information zur Entwicklung in den „kleineren“ Gruppen zur Verfügung stehen. Die Trefferquote viel hier ein wenig besser aus im Vergleich zu dem Ansatz mit KF-Anteilen. Allerdings nicht besser als die ursprüngliche oder die k-Means Gruppierung und man entschied sich bei der ursprünglichen Gemeindegruppierung zu bleiben.

7.2.3 Gradient-Boosting-Verfahren

Zur Modellierung der Nichtanerkennungsquote wurde alternativ versucht, ein geeignetes Modell mittels Gradient-Boosting zu finden. Das Gradientenverfahren wird in der Numerik eingesetzt, um allgemeine Optimierungsprobleme zu lösen. Es basiert auf dem Konzept des sequentiellen Zusammenfügens von Modellen, um die Vorhersagegenauigkeit zu erhöhen. Umgesetzt wurde dieses Modell mittels extreme Gradient Boosting (XGBoost), einer Open-Source Softwarebibliothek für die Programmiersprache R.

Bei der Ergebnis-Betrachtung der richtig, beziehungsweise falsch vorhergesagten Löschungen erkennt man, dass diese Methode in etwa gleich gut ist wie die eine logistische Regression, wobei sie für die Gemeinden mit vielen Klärfällen etwas schlechtere Resultate lieferte, für jene mit wenigen Klärfällen geringfügig bessere.

8 Zusammenfassung

Mit der Registerzählung 2021 steht nun, wie 2011, erneut ein umfangreicher historischer Datenbestand zur Masse der Klärfälle zur Verfügung, mit dem das statistische Modell zur Schätzung der individuellen Löschwahrscheinlichkeit aktualisiert wurde. Dabei wurde ein Standardverfahren – nämlich ein logistisches Regressionsverfahren – angewandt um ein plausibles Modell zu erhalten, welches fortan – zumindest bis zur nächsten Registerzählung bzw. Wohnsitzanalyse - alljährlich im Rahmen der registerbasierten Bevölkerungsfeststellung verwendet werden kann, um die Nichtanerkennungsquoten auf Gemeindeebene zu ermitteln. Das bisher verwendete Modell wurde evaluiert und mit dem Datenbestand der letzten zehn Jahre erweitert. Alternativen zum bisherigen Modell (Gradient-Boosting, alternatives Clustering) wurden verglichen, aber keine signifikante Verbesserung festgestellt.

9 Literaturverzeichnis

- [1] Statistik Österreich: Bericht über die Probezählung 2006
- [2] Bundesgesetz über die Durchführung von Volks-, Arbeitsstätten-, Gebäude- und Wohnungszählungen (Registerzählungsgesetz), BGBl. I Nr. 33/2006 i.d.g.F.
- [3] Statistik Österreich: Berechnungsmethode zur Ermittlung der Nichtanerkennungsquote nach der Registerzählung 2011
- [4] Tom Fawcett: *ROC Graphs: Notes and Practical Considerations for Data Mining Researchers*