

Demografischer Wandel und Wahlverhalten im Schwalm-Eder-Kreis

Moritz Schumacher

Hochschule Hof

Hof / Edermünde, Germany

moritz.schumacher@hof-university.de / moritz.schumacher98@gmail.com

Abstract

Diese Arbeit untersucht den Zusammenhang zwischen demografischem Wandel und lokalem Wahlverhalten in 14 Kommunen des Schwalm-Eder-Kreises (Nordhessen) im Zeitraum 2008 bis 2026. Datengrundlage sind 426 Wahldaten aus dem Votemanager-System der ekom21 sowie 129 demografische Indikatoren des Wegweiser Kommune der Bertelsmann Stiftung. Die Analyse folgt dem CRISP-DM-Prozess und kombiniert K-Means-Clustering, multiple Regressionsanalyse mit Varianzdekomposition, Random-Forest-Klassifikation sowie eine Out-of-Sample-Validierung auf sechs externen Kommunen. Zentraler Befund: Der Wanderungssaldo ist der dominierende demografische Prädiktor – er erklärt 94,4 Prozent der demografisch bedingten AfD-Varianz und 83,6 Prozent der SPD-Varianz, jeweils in entgegengesetzter Richtung. FDP-Ergebnisse hängen primär von der Arbeitsplatzzentralität ab, GRÜNEN-Ergebnisse vom natürlichen Saldo. CDU und FREIE WÄHLER sind demografisch kaum erklärbar. Eine zweistufige Prognose zeigt für 2029 steigende AfD- und sinkende SPD-Werte in allen 14 Kommunen. Die Ergebnisse sind in einem öffentlich zugänglichen interaktiven Dashboard verfügbar.

Demografischer Wandel, Wahlverhalten, Kommunalebene, Wanderungssaldo, CRISP-DM, Random Forest, Regressionsanalyse, Schwalm-Eder-Kreis

I. EINLEITUNG

In vielen Städten und Gemeinden Deutschlands im ländlichen Raum, so auch in Hessen, zeigt sich ein allgemeiner, schleichender Wandel des Bevölkerungsrückgangs. Dazu gehört Abwanderung und Alterung der Menschen in den Städten und Dörfern. Gleichzeitig verändern sich Wahlergebnisse sich drastisch in kurzer Zeit, gleichzeitig können auch kontinuierliche Stimmverluste und -gewinne verschiedener Parteien beobachtet werden. Bisher wurde ein Zusammenhang der Bevölkerungsentwicklung und des Wahlverhaltens auf kommunaler Ebene nicht systematisch untersucht. Dieses Papier soll an dieser Lücke exemplarisch ansetzen. Als Mitglied des Kreisvorstands, Gemeindeverbandsvorsitzender und Gemeindevertreter im Schwalm-Eder-Kreis interessiert mich nicht nur die akademische Fragestellung, ob es einen Zusammenhang von Wahlentwicklung und Demografie gibt, sondern auch eine praktische Motivation: die hessische Kommunalwahl im März 2026 hat deutlich innerhalb des Kreises deutlich gezeigt, dass die Ergebnisse heterogen ausfallen, ohne dass dafür eine systematische und begründbare Erklärung vorhanden wäre.

Die zentrale Forschungsfrage der Arbeit lautet: Welche demografischen Faktoren lokale Wahlunterschiede in den 14 Kommunen im Schwalm-Eder-Kreis im Zeitraum von 2008

bis 2026 erklären? Ziel ist es lokale Erklärungen zu finden, und bundes- oder landespolitische Trends herauszunehmen. Interessant ist hierbei die Lokale Abweichung des lokalen Wahlergebnisses zum übergeordneten Wahlergebnis. Teil der Frage ist es Indikatoren zu finden, warum eine Partei eine lokale Abweichung gegenüber dem übergeordneten Trend aufweist.

Methodisch folgt diese Arbeit dem etablierten Rahmenwerk des CRISP-DM Prozess [1]. Als Datengrundlage dienen maßgeblich zwei öffentlich zugängliche Quellen. Die Wahlergebnisse werden aus dem Votemanager-System (Hersteller ist die votegroup GmbH) der ekom21 [2] bezogen [3]. Die demografischen Daten werden aus dem Wegweiser Kommune der Bertelsmann Stiftung abgerufen [4]. Die Analyseergebnisse fließen in ein interaktives Dashboard ein, dass kommunalpolitisch Interessierten zugänglich gemacht werden soll. Weiter sollen daraus Strategiepapiere für die lokalpolitische Arbeit entstehen.

Die Arbeit ist wie folgt aufgebaut: Abschnitt II ordnet diese Arbeit in den bestehenden Forschungskontext ein. Im Abschnitt III wird das Untersuchungsgebiet und die Forschungsfrage beschrieben. Der Abschnitt IV widmet sich den Datenquellen, die Datenvorbereitung wird hingegen im Abschnitt V dargestellt. Die eingesetzten Modelle zur Analyse werden im Abschnitt VI beleuchtet. Der Abschnitt VII richtet den Fokus auf die Ergebnisse und Evaluation, die im Abschnitt VIII diskutiert werden. Im letzten Abschnitt, Abschnitt IX, wird ein Fazit und Ausblick gegeben.

Für die Erstellung der notwendigen Python-Skripte und Analyseschritte, sowie zur Verfeinerung der Methodik wurde Claude, bzw. Claude Code von Anthropic genutzt. Zur Korrelationsberechnung wurde KNIME verwendet. Der durch mehrfache Iterationen entstandene Workflow ist im Anhang beschrieben (siehe Anhang-Abbildung 1).

II. VERWANDTE ARBEITEN

Auf nationaler Ebene ist der Einfluss des demografischen Wandels bereits erforscht und politikwissenschaftlich untersucht. Kommunale Einflüsse sind jedoch nicht dabei näher wissenschaftlich betrachtet. So haben die Autoren Franz, Fratzscher und Kritikos untersucht, in welchem Umfeld die AfD bei der Bundestagswahl 2017 gut abschnitt. Der Zuspruch für die AfD war in Wahlkreisen relativ hoch, in denen überdurchschnittlich viele Ältere leben – ein Zusammenhang, der in ostdeutschen Wahlkreisen häufiger auftritt als in westdeutschen [5]. In einer Analyse von Bergmann et al. [6], bei der als AfD-Hochburgen identifizierte Landkreise untersucht wurden, konnten unterschiedliche regionalspezifische Muster identifiziert werden. Die MIDEM-Studie [7] zeigt,

dass der positive Zusammenhang zwischen allgemeiner Wanderungsbilanz und rechtspopulistischem Wahlverhalten unabhängig von anderen Faktoren wie Bevölkerungsdichte, BIP pro Kopf oder Arbeitslosenquote gilt. MIDEM untersuchte die Wanderungsbewegungen der vergangenen drei Jahrzehnte auf Ebene von Kreisen und kreisfreien Städten und verglich diese mit dem Ergebnis der Europawahl 2019.

III. PROBLEMVERSTÄNDNIS UND UNTERSUCHUNGSGEBIET

A. Forschungsfrage und Erkenntnisziel

Das Ziel dieser Arbeit ist es, messbare Zusammenhänge zwischen demografischen Indikatoren und dem lokalen Wahlverhalten auf Gemeindeebene zu identifizieren und zu quantifizieren. Lokales Wahlverhalten wird dabei als Wahlresiduum umgesetzt: die Differenz zwischen dem tatsächlichen Ergebnis einer Partei in einer Gemeinde und dem auf Basis des übergeordneten Trends erwarteten Ergebnis. Als übergeordneter Trend werden hier die Wahlergebnisse für alle Wahlberechtigten betrachtet. D.h. bei Bundestagswahlen wird das Bundestagswahlergebnis als Referenz festgehalten, bei Landtagswahlen, das hessische Landtagswahlergebnis. Besonderheit bei Europawahlen ist, dass hier das nationale Ergebnis genutzt wird. Für die Kommunalwahlen wird das Kreiswahlergebnis als Referenz sowohl für die Kreistagswahl als auch die Gemeinde- bzw. Stadtparlamentswahl genutzt. Dieser Ansatz stellt sicher, dass nicht bundespolitische Schwankungen modelliert werden, sondern genuine lokale Abweichungen.

Neben der Erklärung der bisherigen Wahlergebnisse, soll auch eine Prognose für 2029 auf Basis der demografischen Entwicklung für die Parteiergebnisse erzeugt werden. Diese Prognosen dienen nicht als Vorhersagen, sondern als Szenarien, die den kommunalpolitischen Planungs- und Strategieprozess unterstützen sollen.

B. Der Schwalm-Eder-Kreis

Das Untersuchungsgebiet umfasst 14 Kommunen im nordhessischen Schwalm-Eder-Kreis. Dazu gehören die Städte Homberg (Efze), Fritzlar, Melsungen, Gudensberg, Nidenstein, Schwalmstadt, Felsberg, Wabern und Borken (Hessen), sowie die Gemeinden Edermünde, Guxhagen, Neukirchen, Spangenberg und Frielendorf. Untersucht wurden die Wahlergebnisse ab 2008 bis 2026. Berücksichtigt wurden alle Wahlen, die nicht ausschließlich als Personenwahlen – also Landrats- oder Bürgerbürgermeisterwahlen – stattgefunden haben. Bürgerentscheide wurden ebenfalls nicht berücksichtigt. So wurden letztlich Bundes-, Landtags-, und Europawahlen, sowie die Kommunalwahlen für den Landkreis (Kreistag) und die Stadtverordneten- bzw. Gemeindevertretungen berücksichtigt. Die weiteren 13 Kommunen des Landkreises Schwalm-Eder konnten nicht weiter berücksichtigt werden, da der Wegweiser Kommune ausschließlich Kommunen mit einer Einwohnerzahl von mehr als 5.000 Einwohnern berücksichtigt.

Der Schwalm-Eder-Kreis eignet sich als Untersuchungsgebiet, da er auf engem Raum sowohl stagnierende als auch schrumpfende Kommunen vereint. Dieser innerkommunale Vergleich erlaubt es, demografische Unterschiede als Erklärungsfaktor zu isolieren, während geografische und kulturelle Nähe konstant gehalten sind.

IV. DATEN UND DATENVERSTÄNDNIS

A. Wahldaten

Die Wahldaten entstammen dem Votemanager-System von ekom21 [3], dem offiziellen Wahlergebnisinformationssystem des Landes Hessen. Für die 14 untersuchten Kommunen wurden insgesamt 426 Dateien erhoben, davon 284 im HTML-Format und 142 als PDF. Die Dateien decken die oben dargestellten fünf Wahltypen über den gesamten Beobachtungszeitraum ab.

Eine zentrale technische Herausforderung bestand in der Heterogenität des Datenmaterials: Der Votemanager wurde im Laufe des Untersuchungszeitraums mehrfach grundlegend überarbeitet, was drei verschiedene HTML-Generationen erzeugt. Wahlseiten bis etwa 2016 verwenden flache HTML-Tabellen ohne eingebettete Strukturinformationen, neuere Seiten nutzen JSON-Datenbänder mit Summenzeilen, und die jüngsten Seiten ab etwa 2022 sind als Vue.js-Shells realisiert, die keine direkten Tabellendaten enthalten und daher als PDF gespeichert werden mussten. Zusätzlich sind Parteibezeichnungen nicht normiert: Dieselbe Wählergruppe erscheint in verschiedenen Wahlen und Gemeinden unter unterschiedlichen Bezeichnungen. Die verfügbaren CSV-Dateien durch den Votemanager, haben sich als unbrauchbare Datenquelle erwiesen, da hier gar keine Partei und Stimmergebniszuordnung direkt erfolgen. Hierzu wird durch die einzelnen Wahlämter ein separater Text in HTML mit einer Zuordnung angegeben. Nach kurzer Prüfung konnte festgestellt werden, dass abhängig von den Gemeinden überhaupt Texte zu Verfügung gestellt werden. Bei Europawahlen entfielen fast alle Texte.

B. Demografiedaten

Die Demografiedaten stammen aus dem Wegweiser Kommune der Bertelsmann Stiftung [4], die jährlichen Indikatoren auf Gemeindeebene für Deutschland bereitstellt. Die Ausgangsmenge umfasste 165 Indikatoren für den Zeitraum 2006 bis 2023 in sechs Themenfeldern: Bevölkerung, Wirtschaft, Bildung, Soziales, Finanzen und Infrastruktur. 36 Indikatoren wurden wegen unvollständiger zeitlicher Abdeckung ausgeschlossen, definiert als weniger als 20 Prozent vorhandener Beobachtungen. Die verbleibenden 129 Indikatoren bilden die Grundlage der Modellierung. Die Rohdaten lagen im CSV-Format vor, waren jedoch nicht direkt analysierbar: Die Datei verwendet latin-1-Kodierung, Semikolon als Trennzeichen, und die Spaltenköpfe haben das Format 'JAHR\nGEMEINDE' – Jahr und Gemeindegemeinde sind durch einen Zeilenumbruch innerhalb einer einzigen Zelle getrennt. Weiter liegen die einzelnen Indikatoren zeilenweise vor (ein Indikator pro Zeile, eine Spalte pro Gemeinde-Jahr-Kombination). Für den weiteren Gebrauch in der Analyse war jedoch eine Zeile pro Gemeinde und Jahr nötig. Am Ende der Datei lagen weiter noch Fußnoten ohne Messdateninhalt vor, die nicht weiter genutzt und herausgefiltert werden konnten.

Ergänzt wurden die Daten um geografische Koordinaten und Urbanisierungsgrad aus dem Gemeindeverzeichnis des Statistischen Bundesamtes [8].

C. Explorative Datenanalyse

Die explorative Analyse zeigt deutliche Muster in beiden Datensätzen. Die SPD ist in fast allen Kommunen die stärkste Partei, erfährt aber einen kontinuierlichen Rückgang der

Wahlergebnisse. Die CDU stagniert weitestgehend, unter Einbeziehung der Residualanalyse lässt sich ein leichter Aufwärtstrend für die CDU beobachten. Die AfD tritt erst ab 2013 in Erscheinung und gewinnt ab 2017 erheblich an Stärke. Für alle untersuchten Kommunen ist ein Bevölkerungsrückgang erkennbar. 11 von 14 Kommunen haben einen negativen Wanderungssaldo. Eine erste visuelle Inspektion zeigt einen positiven Zusammenhang zwischen negativem Wanderungssaldo und erhöhtem AfD-Residuum, was die Ausgangshypothese stützt.

V. DATENVORBEREITUNG

A. Wahldaten: Parsing, Normalisierung und Datenbankaufbau

Zur automatisierten Verarbeitung der 426 heterogenen Quelldateien wurde ein Python-Parser entwickelt, der alle drei HTML-Generationen sowie PDF-Dateien unterstützt. Je nach erkanntem Dateityp kommen unterschiedliche Parsing-Strategien zum Einsatz: Für neue HTML-Dateien werden Summenzeilen aus strukturierten Tabellen extrahiert, für alte HTML-Dateien wird eine flache Tabelle einzeilig eingelesen, und für PDFs erfolgt eine regelbasierte Textextraktion mit regulären Ausdrücken. Vue.js-Shells, die keine eingebetteten Daten enthalten, werden als Fallback markiert und erfordern eine manuelle PDF-Version.

Ein eigener Normalisierungsschritt überführt alle Partei-bezeichnungen in ein kanonisches Schema. So werden FWG, FWG Borken, FWGF und verwandte Bezeichnungen einheitlich als FREIE WÄHLER erfasst, PDS und Linkspartei als DIE LINKE. Die bereinigten Daten werden in einer SQLite-Datenbank mit drei Tabellen gespeichert: election (Metadaten je Wahl), statistic (Wahlstatistik) und result (Parteiergebnisse je Zeile). Eine automatische Plausibilitätsprüfung vergleicht die Summenstimmen mit der ausgewiesenen Zahl gültiger Stimmen.

B. Demografiedaten: Formatkonvertierung und Normalisierung

Die Aufbereitung der Wegweiser-Kommune-Daten erforderte mehrere Transformationsschritte, die in einem eigenen Python-Skript implementiert wurden. Im ersten Schritt werden die Spaltenköpfe geparkt: Der Zeilenumbruch innerhalb jeder Zelle wird erkannt und in Jahr und Gemeindefname aufgetrennt. Da die Gemeindefnamen im Wegweiser nicht mit den Gemeindefkürzeln in der Wahldatenbank übereinstimmen – der Wegweiser verwendet zum Beispiel 'Borken (Hessen)', die Wahldatenbank hingegen 'borken'. Ein manuelles Mapping-Wörterbuch wurde entsprechend angewendet, das alle 14 Kommunen eindeutig zuordnet.

Im zweiten Schritt wird das Format transponiert: Aus der Ausgangsstruktur mit Indikatoren als Zeilen und Gemeinde-Jahr-Kombinationen als Spalten entsteht eine Tabelle, in der jede Zeile eine Gemeinde zu einem Jahr repräsentiert und jede Spalte einem Indikator entspricht. Fußnoten-Zeilen am Ende der CSV-Datei, die keine Messwerte enthalten, werden vor der Transposition herausgefiltert. Anschließend werden 36 Indikatoren wegen unvollständiger zeitlicher Abdeckung (wie oben bereits definiert) ausgeschlossen. Die verbleibenden 129 Indikatoren werden durch geografische Koordinaten und den Urbanisierungsgrad aus dem Gemeindeverzeichnis des Statistischen Bundesamtes ergänzt.

C. Zeitversatzverknüpfung und Residualanalyse

Für die Verknüpfung von Wahl- und Demografiedaten wird ein Zeitversatz von einem Jahr verwendet: Demografiedaten zum Zeitpunkt t-1 werden mit dem Wahlergebnis zum Zeitpunkt t verbunden. Es wird angenommen, dass demografische Entwicklungen mit einer gewissen Verzögerung in Wahlentscheidungen sichtbar werden.

Im Anschluss daran, wird das Residuum für jede Partei und Gemeinde berechnet. Als Referenz wird wie oben angegeben, das Wahlergebnis im gesamten Wahlgebiet genutzt. Die Differenz zwischen Gemeindergebnis und gesamten Wahlergebnis ist die Zielvariable der Modellierung. Ohne diesen Schritt würden im weiteren Verlauf übergreifende politische Trends auf lokale Entwicklungen übertragen werden – ohne dass eine Verknüpfung bestehen muss, kommunale Spezifika könnten so nicht abgebildet werden. Die Residualanalyse ist damit ein zentrales methodisches Element dieser Untersuchung.

VI. MODELLIERUNG

A. Demografische Clusteranalyse

Um verschiedene Kommuntypen zu identifizieren, wurde mit Hilfe von K-Means [9] zwei Cluster von Gemeindetypen auf Basis der 129 demografischen Indikatoren ermittelt. Dabei wurde für den Typ A – Kommunen mit starker Erwerbsbasis, geringerem Altersdurchschnitt, mit positivem oder ausgeglichenem Wanderungssaldo – die Gemeinden bzw. Städte Guxhagen, Edermünde, Homberg, Melsungen, Fritzlar, Wabern, Niedenstein und Gudensberg zugeordnet. Der Typ B umfasst die Kommunen Schwalmstadt, Spangenberg, Neukirchen, Borken, Frielendorf und Felsberg. Dargestellt werden diese in der Abbildung 1. Diese Kommunen weisen einen deutlichen Bevölkerungsrückgang, höheren Altersdurchschnitt und negativen Wanderungssaldo aus. Typ B-Kommunen weisen einen systematisch höheren AfD-Residuen aus – die SPD hingegen hat hier niedrigere Residuen auszuweisen. Typ A-Kommunen zeigen eher positive Residuen für die SPD aus. Eine wesentliche Ausnahme stellt hier Fritzlar dar – in Fritzlar sind die Residuen der CDU deutlich gegenüber den Kommunen gleichen Typs erhöht.

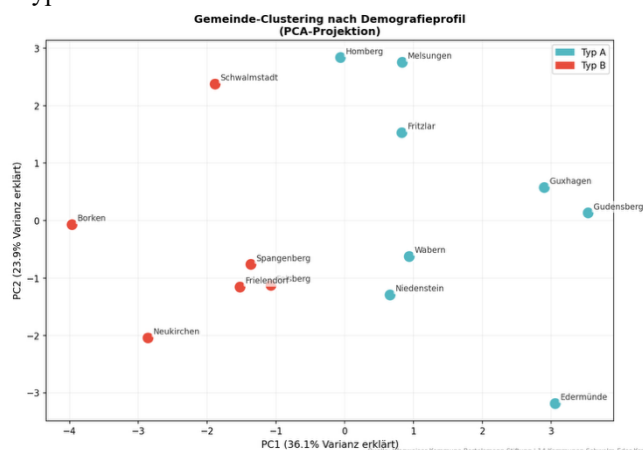


Abbildung 1

B. Multiple Regressionsanalyse

Für die Regressionsanalyse [10] der Daten wurden insgesamt sechs Regressionsmodelle für jede Partei

durchgeführt. Es wurden die sechs größten Parteien für die Analyse ausgewählt. Es wurden folglich CDU, SPD, AfD, Bündnis 90/Grüne (nachfolgend nur GRÜNE), FDP und FREIE WÄHLER untersucht. Kleinere Parteien oder örtliche Wählerlisten treten nicht bei allen untersuchten Wahltypen an und sind daher hier nicht weiter betrachtet worden. Die Partei "DIE LINKE" die Bundesweit vertreten ist [11], hat aber in allen 14 untersuchten Kommunen keine absoluten Stimmanteile größer als drei Prozent, weshalb eine belastbare Regressionsanalyse nicht möglich war. Daher wurde diese Partei aus der Modellierungsphase ausgeschlossen.

Zur Sicherstellung stabiler Koeffizientenschätzungen wurde für jedes Modell zunächst eine VIF-Kontrolle¹ durchgeführt, gefolgt von einer schrittweisen Rückwärtselimination nicht-signifikanter Prädiktoren. In der Tabelle I werden die Ergebnisse zentral zusammengefasst. Mit einem Bestimmtheitsmaß von $R^2 = 0,38$ weist das AfD-Modell den stärksten demografischen Erklärungsgehalt aller sechs untersuchten Parteien auf. Dies deutet darauf hin, dass AfD-Wahlergebnisse im Schwalm-Eder-Kreis stärker mit demografischen Veränderungen zusammenhängen als die Ergebnisse der übrigen Parteien, wenngleich 62 Prozent der Varianz durch nicht erfasste Faktoren bedingt bleibt. Stärkster Prädiktor für die Entwicklung der Residuen der Parteien ist hier das Wanderungssaldo. Darüber hinaus sind in unterschiedlichem Maß weitere demografische Faktoren relevant: Für die GRÜNEN dominiert der natürliche Saldo, für die FDP die Arbeitsplatzzentralität.

Die Varianzdekomposition zeigt, dass die erklärte Varianz in den einzelnen Modellen auf wenige – teils nur einen einzigen – demografischen Faktor konzentriert ist. Beim AfD-Modell entfallen 94,4 Prozent des erklärten Varianzanteils auf den Wanderungssaldo². Der verbleibende Anteil von 5,6 Prozent erklärt sich durch den natürlichen Saldo.

Das AfD-Wahlverhalten im Schwalm-Eder-Kreis ist damit in dem Maße, wie es demografisch erklärbar ist, nahezu vollständig durch die Netto-Migration bestimmt. Ein vergleichbares Muster zeigt sich beim SPD-Modell: 83,6 Prozent des erklärten Varianzanteils entfallen auf den Wanderungssaldo, die verbleibenden Anteile verteilen sich auf Arbeitsplatzzentralität (9,5 Prozent) und natürlichen Saldo (7,0 Prozent). AfD und SPD reagieren damit auf denselben demografischen Faktor – allerdings in entgegengesetzte Richtungen. Kommunen mit negativem Wanderungssaldo weisen höhere AfD-Residuen und niedrigere SPD-Residuen auf. Für FDP und GRÜNE zeigen sich abweichende Muster. Beim FDP-Modell dominiert die Arbeitsplatzzentralität mit 85,1 Prozent des erklärten Varianzanteils: Kommunen mit starker Erwerbsanbindung an wirtschaftliche Zentren weisen systematisch höhere FDP-Residuen auf.

Bei den GRÜNEN tritt der natürliche Saldo in den Vordergrund (73,8 Prozent), während der Wanderungssaldo nur 26,2 Prozent beisteuert. CDU und FREIE WÄHLER weisen

mit R^2 -Werten von 0,05 bzw. 0,04 insgesamt sehr geringe Bestimmtheitsmaße auf. Eine inhaltliche Interpretation der Varianzanteile ist daher nur eingeschränkt belastbar.

Tabelle I

Partei	R^2	Stärkster Prädiktor	Accuracy (RF auto) ^a	MAE ³ Validierung in PP ^b
AfD	0,38	Wanderungssaldo	0,69	4,02
SPD	0,27	Wanderungssaldo	0,79	8,78
GRÜNE	0,21	Bildungsniveau	0,69	2,53
FDP	0,18	Arbeitsplatzzentral.	0,83	2,22
CDU	0,09	(kein signif.)	0,77	5,47
FW ^c	0,04	(kein signif.)	0,61	6,00

^aRF = Random Forest; ^bPP = Prozentpunkte; ^cFW = FREIE WÄHLER

C. Klassifikation mittels Random Forest

Zur Validierung der Regressionsresultate wurde ergänzend ein Random-Forest-Klassifikationsmodell eingesetzt [12]. Random Forest eignet sich für diesen Zweck, da es als nicht-lineares Verfahren keine Annahmen über die Verteilung der Daten trifft und damit unabhängig von den Voraussetzungen der linearen Regression arbeitet. Übereinstimmungen zwischen beiden Methoden erhöhen die Belastbarkeit der Befunde.

Als Zielvariable diente – analog zur Regressionsanalyse – das demografische Residuum je Gemeinde und Partei. Um die Feature-Auswahl methodisch konsistent mit der Regressionsanalyse zu gestalten, wurde anstelle einer manuellen

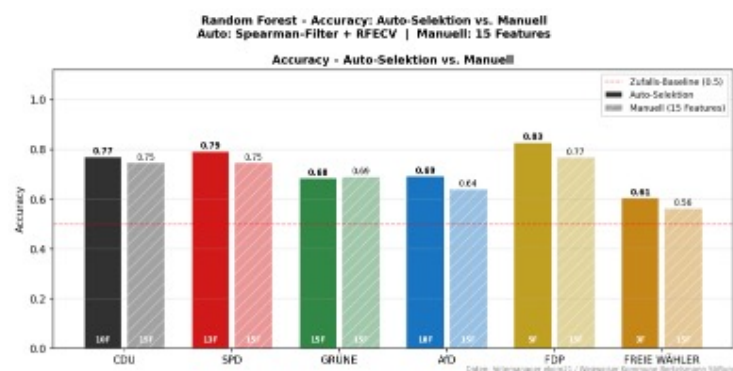


Abbildung 2

¹ Der Variance Inflation Factor (VIF) misst das Ausmaß der Multikollinearität zwischen Prädiktoren eines Regressionsmodells. Er gibt an, um welchen Faktor die Varianz eines Koeffizienten durch Korrelation mit anderen Prädiktoren aufgebläht wird. Ein VIF-Wert über 10 gilt als Schwellenwert für problematische Kollinearität; betroffene Variablen werden vor der Modellschätzung ausgeschlossen, um stabile Koeffizientenschätzungen sicherzustellen [10].

² Die Stichprobengröße der AfD (n=106) ist kleiner als die der anderen Parteien, da die AfD erst ab 2013 antritt.

³ MAE – Mean Absolute Error: Er gibt an, um wie viele Prozentpunkte die Modellvorhersage im Durchschnitt vom tatsächlichen Wahlergebnis abweicht.

Vorauswahl eine zweistufige automatische Selektion eingesetzt: In einem ersten Schritt wurden hochkorrelierte Features durch einen Spearman-Korrelationsfilter (Schwellenwert: $r > 0,85$) bereinigt – analog zur VIF-Kontrolle in der Regressionsanalyse. Im zweiten Schritt erfolgte eine Recursive Feature Elimination with Cross-Validation (RFECV), die iterativ die optimale Featuremenge per fünffacher Kreuzvalidierung bestimmt – analog zur Rückwärtselimination. Die resultierende Featuremenge variiert je Partei zwischen drei Features (FREIE WÄHLER) und 18 Features (AfD).

Die Klassifikationsgenauigkeit (Accuracy) der automatisch selektierten Modelle ist in Tabelle I ausgewiesen. Die FDP erreicht mit 0,83 den höchsten Wert des gesamten Datensatzes. Fünf Features genügen dem Modell, um FDP-Hochburgen zuverlässig zu identifizieren. Arbeitsplatzzentralität und Gewerbesteuer dominieren dabei deutlich. SPD (0,79) und CDU (0,77) folgen. AfD und GRÜNE liegen bei 0,69, FREIE WÄHLER bei 0,61. Im Vergleich zur manuellen 15-Feature-Variante verbessern sich alle Parteien außer den GRÜNEN, bei denen beide Varianten identische F1-Werte erzielen.

Die Feature-Importance-Analyse zeigt teilweise abweichende Ergebnisse zur Regressionsanalyse. Für die FDP und CDU bestätigt der Random Forest die Dominanz der Arbeitsplatzzentralität. Bei der AfD identifiziert der Random Forest jedoch „Haushalte mit niedrigem Einkommen“ und „Familieneinwanderung“ als wichtigste Prädiktoren. Das Wanderungssaldo, aus der Regressionsanalyse mit 94,4 Prozent der erklärten Varianz trägt, hingegen nicht. Der Random Forest erfasst damit einen anderen Aspekt der Prekaritätsdimension und deutet darauf hin, dass sozioökonomische Verdrängungseffekte und selektive Abwanderung von Familien eigenständige Erklärungskraft für AfD-Wahlergebnisse besitzen, die von der Regressionsanalyse im Wanderungssaldo aggregiert werden. Bei der SPD tritt „Haushalte mit Kindern“ als wichtigster Prädiktor in den Vordergrund, was den demografischen Strukturwandel in familiengeprägten Kommunen als eigenständigen Erklärungsfaktor unterstreicht. Bei den GRÜNEN bestätigt der Random Forest die Dominanz bildungsnaher Indikatoren – „Hochqualifizierte am Wohnort“ führt die Feature Importance an.

Diese Rangfolge stimmt in den Grundzügen mit den Ergebnissen der Varianzdekomposition aus Abschnitt VI-B überein. Abweichungen bei AfD und SPD sind kein Widerspruch, sondern eher ein Hinweis darauf, dass das Wanderungssaldo und sozioökonomische Prekarität eng verwobene Dimensionen darstellen, die je nach Modellarchitektur unterschiedlich aufgelöst werden. Die gesamte Darstellung der Feature-Importance ist in Anhang-Abbildung 2 sichtbar.

D. Prognose 2029

1) Prognose auf Basis der Regressionsanalyse

Für die Wahlprognose 2029 wurde ein zweistufiges strukturelles Prognosemodell entwickelt, das auf in der Wahlforschung etablierten Fundamentalansätzen basiert

[13]. In der ersten Stufe wurde für jede der sechs Parteien und jede der 14 Kommunen eine lineare Trendregression auf die historischen Wahlergebnisse gelegt. Dieses Vorgehen entspricht einer Status-quo-Extrapolation: Es wird angenommen, dass sich die in der Vergangenheit beobachteten lokalen Entwicklungen unter Abwesenheit struktureller Brüche fortschreiben lassen [14]. In der zweiten Stufe wurden die demografischen Projektionswerte des Wegweiser Kommune für 2029 in die bestehenden Regressionsmodelle eingespeist, um das demografisch bedingte Residuum für das Prognosejahr zu berechnen. Der Prognosewert ergibt sich als Summe aus lokalem Trend und demografischem Residuum. Konfidenzintervalle wurden auf Basis der Standardfehler der Regressionsmodelle berechnet. Die Prognose erhebt keinen Anspruch auf punktgenaue Vorhersage, sondern zeigt Entwicklungsszenarien unter der Annahme konstanter struktureller Bedingungen auf.

Das Modell zeigt für die AfD in allen 14 Kommunen eine positive Trendentwicklung. Die Bandbreite der Prognosewerte reicht von 17,8 Prozent in Guxhagen bis zu 45,2 Prozent in Felsberg. Der Felsberg-Wert ist dabei mit Vorsicht zu interpretieren, da ab 2018 keine vollständigen Wahldaten vorliegen und der Trendkoeffizient auf einem verkürzten Beobachtungszeitraum basiert.⁴ In den übrigen Kommunen liegen die AfD-Prognosewerte zwischen 20 und 34 Prozent und sind damit belastbarer. Der demografische Anteil am Prognosewert – also der durch den Wanderungssaldo erklärte Zusatzimpuls – beträgt im Schnitt 2,1 Prozentpunkte und verstärkt damit den positiven Trendanteil systematisch. Für einen stärkeren positiven Zusatztrend steht Niedenstein, beispielhaft dargestellt in Anhang-Abbildung 3.

Für die SPD zeigt das Modell in allen 14 Kommunen eine negative Entwicklung. Die Prognosewerte reichen von 17,4 Prozent in Niedenstein bis 31,4 Prozent in Edermünde. In Felsberg überlagern sich negativer lokaler Trend (–0,85 PP pro Jahr) und ein stark negativer demografischer Impuls von –11,8 Prozentpunkten zu einem Prognosewert von 19,2 Prozent – dem stärksten absoluten Rückgang im gesamten Datensatz. Für die CDU ergibt das Modell moderate Verluste von durchschnittlich zwei bis drei Prozentpunkten gegenüber dem Wert von 2023, mit Ausnahme von Felsberg (–6,5 PP). Die GRÜNEN zeigen eine weitgehend stabile Entwicklung mit leichtem Aufwärtstrend in bevölkerungsstabileren Kommunen (Typ A-Kommunen, siehe Abschnitt VI-A). Für FDP und FREIE WÄHLER sind die Konfidenzintervalle aufgrund der niedrigen Bestimmtheitsmaße sehr breit; eine belastbare Punktprognose ist für diese Parteien nicht möglich. Tabelle II zeigt die stärksten prognostizierten Veränderungen gegenüber dem Ergebnis von 2023.

2) Prognose auf Basis des Random Forest und Methodenvergleich

Zur Überprüfung der regressionsbasierten Prognose wurde das Random-Forest-Modell auf denselben demografischen Projektionswerten für 2029 angewendet. Beide Methoden wurden auf identischer Datengrundlage betrieben, was einen direkten Methodenvergleich ermöglicht.

⁴ Für Felsberg liegen ab 2018 keine vollständigen Wahldaten vor; der Prognosewert basiert auf einem verkürzten Beobachtungszeitraum und ist mit erhöhter Unsicherheit behaftet (KI: 38,6–51,7%).

Als Konsistenzkriterium gilt eine maximale Abweichung von zwei Prozentpunkten zwischen beiden Prognosen.

Die Kreuzvalidierungsgüte ($CV-R^2$) des Random-Forest-Modells – also seine Fähigkeit, auf ungesehenen Daten korrekte Prognosen zu liefern – variiert erheblich zwischen den Parteien. Bei der fünffachen Kreuzvalidierung wird das Modell wiederholt auf einem Teil der Daten trainiert und auf dem zurückgehaltenen Rest getestet. Die GRÜNEN erreichen mit $CV-R^2=0,26$ die höchste Vorhersagegüte aller Parteien, gefolgt von FDP (0,28) und AfD (0,32). Für CDU (–0,13) und SPD (–0,22) sind die Kreuzvalidierungswerte negativ. Das Modell liegt auf ungesehenen Daten schlechter als eine reine Mittelwertschätzung. Für SPD ist das besonders aufschlussreich, da die Klassifikations-Accuracy des Modells mit 0,79 hoch ist: Das Modell erkennt zuverlässig, ob eine Gemeinde überdurchschnittlich oder unterdurchschnittlich abscheidet, kann aber den konkreten Prozentwert nicht belastbar schätzen. SPD-Prognosewerte aus dem Random Forest sind daher als Trendindikator zu verstehen, nicht als Punktprognose.

Die stärkste Übereinstimmung zwischen Regressions- und Random-Forest-Prognose zeigen die GRÜNEN – sowohl methodisch ($CV-R^2=0,26$) als auch inhaltlich: Die Prognosewerte beider Methoden liegen für 7 von 14 Kommunen innerhalb des Konsistenzkriteriums von zwei Prozentpunkten. Dass ein lineares und ein nicht-lineares Verfahren hier zu nahezu identischen Ergebnissen kommen, ist ein methodisch belastbarer Befund und deutet darauf hin, dass der demografische Zusammenhang bei den GRÜNEN tatsächlich weitgehend linear ist.

Tabelle II

Gemeinde	Partei	Aktuell 2023 ^d	Prognose 2029 ^d	Veränderung ^e	Demo-An teil ^e
Frielen-dorf	AfD	24,1	34,1	+10,0	+2,8
Homberg	AfD	22,2	31,4	+9,2	+2,3
Schwalm-stadt	AfD	19,7	28,7	+9,0	+2,0
Nieden-stein	SPD	27,2	17,4	-9,8	-3,9
Borken	SPD	27,5	22,1	-5,5	-1,4
Schwalm-stadt	SPD	25,4	19,0	-6,4	-2,0
Guxhagen	CDU	20,4	17,1	-3,3	-0,5

^d in Prozent; ^e in Prozentpunkten

Die stärkste Abweichung zeigt sich weiterhin bei der AfD. Die Regressionsanalyse prognostiziert Werte zwischen 18 und 34 Prozent, der Random Forest konsistent niedrigere Werte zwischen 14 und 23 Prozent. Diese Differenz ist nach Ausschluss von Felsberg mit einem MAE von 7,9 Prozentpunkten zwischen den Methoden stabil. Der Random Forest

bildet als nicht-lineares Verfahren implizit Sättigungseffekte ab. Bei einem gegebenen Wanderungssaldo gibt es ein begrenztes AfD-Potenzial, das die lineare Regression nicht berücksichtigt. Der plausible Prognosebereich liegt damit bei 15 bis 23 Prozent: Schwalm-Eder-Kommunen lagen historisch 5 bis 10 Prozentpunkte über dem Bundesschnitt, was bei einem bundesweiten AfD-Wert von rund 20 Prozent im Jahr 2029 diesen Bereich stützt.

Insgesamt gilt: Wo beide Methoden in Richtung und Größenordnung übereinstimmen, insbesondere bei den GRÜNEN und der FDP, ist die Prognose methodisch robust. Die systematische Abweichung bei der AfD ist ein Hinweis auf die Grenzen linearer Extrapolation bei nicht-linearen Wachstumsprozessen. Für CDU und SPD empfiehlt sich die Interpretation der RF-Prognose als Richtungsindikator ohne Anspruch auf Punktgenauigkeit. Der Methodenvergleich je Partei und Kommune ist in Anhang-Abbildung 4 ersichtlich.

Tabelle III

Partei	Regression ^e	RF ^a ohne Geo ^e	RF ^a mit Geo ^e	Beste Methode
AfD	4,02	4,52	3,59	RF ^a mit Geo
SPD	8,78	9,17	9,42	Regression
GRÜNE	2,53	2,70	2,67	Regression
FDP	2,22	2,52	2,26	Regression
CDU	5,47	5,94	5,58	Regression
FW ^c	6,00	7,45	6,87	Regression

^aRF = Random Forest; ^cFW = FREIE WÄHLER; ^ein Prozentpunkten

VII. EVALUATION UND VALIDIERUNG

Zur Überprüfung der Generalisierbarkeit wurden alle drei Modellvarianten – Regressionsmodelle, Random Forest ohne Geodaten und Random Forest mit Geodaten [8] – auf sechs externen Kommunen validiert, die nicht Teil der Modellentwicklung waren: Alsfeld, Mücke und Schlitz aus dem Vogelsbergkreis sowie Grebenstein, Naumburg und Schauenburg aus dem Landkreis Kassel. Die Validierung erstreckt sich über alle verfügbaren Wahljahre je Kommune (2017, 2018, 2021 und 2023), was eine robustere Beurteilung erlaubt als eine Validierung auf einem einzelnen Wahltermin. Alle drei Varianten wurden auf identischer Datenbasis berechnet, um eine direkte Vergleichbarkeit sicherzustellen.

Als Gütemaß dient wieder der mittlere absolute Fehler (MAE). Der Methodenvergleich je Partei und Kommune ist in Anhang-Abbildung 4 ersichtlich.

fasst die Ergebnisse zusammen.

Für fünf der sechs Parteien erzielt die Regressionsanalyse den niedrigsten MAE. Die deutlichste Überlegenheit zeigt sich bei FDP (2,22 PP) und GRÜNE (2,53 PP), wo die Regression den RF um 0,26 bis 0,30 Prozentpunkte unterbietet. Diese Ergebnisse bestätigen, dass der demografische Zusammenhang bei diesen Parteien hinreichend linear ist, um

durch Regressionsmodelle auch auf externe Kommunen übertragen zu werden.

Die einzige Ausnahme bildet die AfD: Hier erzielt der Random Forest mit Geodaten den niedrigsten MAE von 3,59 Prozentpunkten gegenüber 4,02 Prozentpunkten der Regression. Der Vorteil ist moderat, aber konsistent. Er deutet darauf hin, dass die nicht-lineare Modellarchitektur des Random Forest die AfD-Muster in Verbindung mit geografischer Kontextinformation besser auf externe Kommunen überträgt. Geodaten verbessern den RF bei der AfD um 0,93 Prozentpunkte – der stärkste Geo-Effekt aller Parteien. Für alle anderen Parteien ist der Geo-Effekt gering oder neutral.

Für die SPD erreichen alle drei Methoden hohe Fehlerwerte zwischen 8,78 und 9,42 Prozentpunkten. Das ist konsistent mit dem negativen CV- R^2 des RF-Regressionsmodells für die SPD und dem niedrigsten R^2 -Wert der Regressionsanalyse: Regionale politische Traditionen, die bei der SPD in diesen Kommunen besonders ausgeprägt sind, entziehen sich systematisch sowohl demografischer als auch geografischer Modellierung.

Insgesamt bestätigt die Validierung die Regressionsanalyse als robuste Standardmethode für externe Kommunen. Der Random Forest mit Geodaten stellt eine überlegene Alternative speziell für die AfD dar.

VIII. DISKUSSION

A. Ergebnisinterpretation und Handlungsempfehlung

Die Analyseergebnisse zeigen, dass demografischer Wandel das lokale Wahlverhalten im Schwalm-Eder-Kreis messbar beeinflusst – allerdings in unterschiedlichem Ausmaß je Partei und mit deutlichen Grenzen.

Der Wanderungssaldo ist der entscheidende demografische Faktor. Auf ihn entfallen 94,4 Prozent der erklärten AfD-Varianz und 83,6 Prozent der erklärten SPD-Varianz. AfD und SPD reagieren dabei auf denselben Indikator in entgegengesetzte Richtungen: In Kommunen mit negativem Wanderungssaldo steigen die AfD-Residuen, die SPD-Residuen fallen. Dieser Befund stimmt mit der MIDEM-Studie [7] überein, die denselben Zusammenhang auf Kreisebene gezeigt hat – er lässt sich hier auf Gemeindeebene und unter Kontrolle des Bundestrends bestätigen.

FDP und GRÜNE folgen anderen Mustern. Für die FDP erklärt die Arbeitsplatzzentralität 85,1 Prozent der Varianz – wirtschaftlich gut angebundene Kommunen weisen höhere FDP-Residuen auf. Für die GRÜNEN dominiert der natürliche Saldo (73,8 Prozent). CDU und FREIE WÄHLER sind durch demografische Indikatoren kaum erklärbar ($R^2 = 0,05$ bzw. $0,04$) – hier spielen lokale Faktoren wie Kandidatenpersönlichkeiten und kommunalpolitische Themen offenbar eine entscheidendere Rolle. Das zeigt sich besonders in Fritzlar, dessen CDU-Residuen auch im Vergleich mit strukturell ähnlichen Typ-A-Kommunen deutlich erhöht sind.

Die Prognosemodelle verlängern diese Muster bis 2029: Die AfD steigt in allen 14 Kommunen, die SPD fällt in allen 14. Der demografische Anteil am AfD-Anstieg beträgt im Schnitt 2,1 Prozentpunkte – moderat, aber systematisch in dieselbe Richtung wie der Trendanteil. Regression und Random Forest stimmen in der Trendrichtung überein; der Random Forest prognostiziert für die AfD konservativere Absolutwerte und bildet damit implizit Sättigungseffekte ab.

Kommunen mit negativem Wanderungssaldo, besonders die Typ-B-Kommunen Schwalmstadt, Borken und Frielendorf, sollten Maßnahmen zur Bevölkerungsstabilisierung priorisieren, da der Wanderungssaldo der stärkste demografische Hebel auf das Wahlverhalten ist. Für Parteien erlauben die Ergebnisse eine gezieltere Ressourcenplanung: FDP-Potenziale liegen strukturell in wirtschaftlich gut angeordneten Kommunen, GRÜNEN-Potenziale in Gemeinden mit günstiger natürlicher Bevölkerungsentwicklung. Die eingesetzte Methodik ist zudem direkt auf andere Landkreise übertragbar, da alle Datenquellen für die Bundesländer Bayern, Baden-Württemberg, Sachsen, Hessen, Nordrhein-Westfalen und Andere verfügbar sind.

Einschränkend ist festzuhalten: R^2 -Werte zwischen 0,05 und 0,38 zeigen, dass Demografie das lokale Wahlverhalten zwar messbar, aber nicht vollständig erklärt. Die Befunde beschreiben Aggregate auf Gemeindeebene. Rückschlüsse auf individuelles Wahlverhalten sind daraus nicht zulässig. Diese und weitere Limitierungen werden im folgenden Abschnitt aufgeführt.

B. Limitierungen

Die Stichprobengröße von 14 Kommunen ist die zentrale Einschränkung dieser Arbeit. Statistisch belastbare Aussagen sind bei dieser Datenbasis nur begrenzt möglich. Overfitting bleibt trotz Kreuzvalidierung ein Risiko. Eine Erweiterung auf alle Votemanager-Kommunen in Hessen oder Deutschland würde die Aussagekraft erheblich steigern.

Darüber hinaus sind zentrale Einflussfaktoren nicht erfasst: Kandidateneffekte, lokale politische Ereignisse, Vereinsstrukturen, religiöse Zugehörigkeit und die regionale Medienlandschaft spielen nachweislich eine Rolle – wie das Beispiel Fritzlar zeigt –, lassen sich aber nicht in demografischen Indikatoren abbilden.

Die Demografiedaten des Wegweiser Kommune basieren auf amtlichen Meldedaten, die mit Verzögerung erfasst werden. Bevölkerungsbewegungen des laufenden Jahres sind darin nicht enthalten, was die Aktualität der Prognosen einschränkt.

Die Prognosen für 2029 beruhen auf linearer Trendextrapolation und setzen konstante strukturelle Bedingungen voraus. Strukturelle Brüche, etwa durch bundespolitische Krisen, Migrationsereignisse oder wirtschaftliche Schocks, werden nicht berücksichtigt. Die Prognosen sind daher als Szenarien zu verstehen, nicht als Vorhersagen.

Schließlich gilt für alle Befunde dieser Arbeit: Da Gemeindedaten analysiert werden, können keine Schlüsse auf individuelles Wahlverhalten gezogen werden. Der ökologische Fehlschluss, also die Übertragung von Aggregatbefunden auf Einzelpersonen, ist eine grundlegende methodische Grenze jeder Analyse auf Gemeindeebene.

C. Praxistransfer: interaktives Dashboard

Die Ergebnisse dieser Arbeit sind in einem interaktiven Dashboard unter cduedermuende.de öffentlich zugänglich. Ziel ist es, die Analyseergebnisse auch ohne datenwissenschaftlichen Hintergrund nutzbar zu machen – für Kommunalpolitiker, Wahlkampfteams und Kommunalverwaltungen im Schwalm-Eder-Kreis.

Das Dashboard visualisiert die Cluster-Zuordnung der 14 Kommunen, die Feature-Importance-Ergebnisse je Partei sowie die Prognosewerte für 2029 mit Konfidenzintervallen.

Nutzer können einzelne Gemeinden und Parteien auswählen und die zugrundeliegenden demografischen Indikatoren einsehen.

Die Veröffentlichung folgt dem CRISP-DM-Deployment-Schritt [1]: Die analytischen Ergebnisse werden in eine Form überführt, die im kommunalpolitischen Planungsprozess direkt einsetzbar ist. Das Dashboard wird mit den Ergebnissen der Bundestagswahl 2029 aktualisiert, sobald diese vorliegen.

Diese Arbeit hat gezeigt, dass demografischer Wandel das lokale Wahlverhalten im Schwalm-Eder-Kreis messbar beeinflusst. Der Wanderungssaldo ist der entscheidende Einzelfaktor: Er erklärt nahezu die gesamte demografisch bedingte Varianz bei AfD und SPD – in entgegengesetzte Richtungen. FDP und GRÜNE folgen anderen demografischen Logiken, CDU und FREIE WÄHLER sind durch Demografiedaten kaum erklärbar.

IX. FAZIT UND AUSBLICK

Der eingesetzte CRISP-DM-Prozess [1] hat sich als geeigneter Rahmen für diese explorative Analyse bewährt. Die Kombination aus fünf Methoden – Clusteranalyse, multiple Regression mit Varianzdekomposition, Random-Forest-Klassifikation, zweistufige Prognose und Out-of-Sample-Validierung – erlaubt es, Befunde methodisch abzusichern und gegeneinander zu prüfen. Wo Regression und Random Forest in Richtung und Größenordnung übereinstimmen, sind die Ergebnisse besonders belastbar.

Die Methodik ist vollständig reproduzierbar und auf andere Landkreise übertragbar. Votemanager und Wegweiser Kommune stehen weitflächig zur Verfügung. Eine Skalierung auf einen Großteil deutscher Kommunen wäre damit technisch möglich und wissenschaftlich reizvoll. Eine solche Erweiterung, verbunden mit einer theoretischen Einbettung in soziologische Erklärungsansätze demografischen Wandels, wäre ein naheliegender Gegenstand für eine weiterführende Masterarbeit.

Das Dashboard unter cduedermuende.de wird nach der Bundestagswahl 2029 mit aktuellen Ergebnissen aktualisiert. Die Prognosewerte können dann erstmals empirisch überprüft werden.

X. LITERATURVERZEICHNIS

- [1] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer und R. Wirth, CRISP-DM 1.0, The CRISP-DM consortium, 2000.
- [2] ekom21, „www.ekom21.de“, ekom21 GmbH, [Online]. Available: <https://www.ekom21.de/loesungen/votemanager/>. [Zugriff am 30 05 2026].
- [3] votegroup GmbH, „wahlen.votemanager.de“, votegroup GmbH, [Online]. Available: <https://wahlen.votemanager.de>. [Zugriff am 15 04 2026].
- [4] Bertelsmann Stiftung, „Wegweiser Kommune“, Bertelsmann Stiftung, [Online]. Available: <https://www.wegweiser-kommune.de>. [Zugriff am 15 04 2026].

- [5] C. Franz, M. Fratzscher und A. S. Kritikos, „AfD in dünn besiedelten Räumen mit Überalterungsproblemen stärker“, *DIW Wochenbericht* Nr. 8, 2018.
- [6] K. Bergmann, M. Diermeier und J. Niehues, „Ein komplexes Gebilde. Eine sozio-ökonomische Analyse des Ergebnisses der AfD bei der Bundestagswahl 2017“, *Zeitschrift für Parlamentsfragen*, Nr. No. 2, p. 243–264, 2018.
- [7] TU Dresden Institut für Politikwissenschaft/ Zentrum für Verfassungs- und Demokratieforschung Philosophische Fakultät, „Emigration in Europa“, Mercator Forum Migration und Demokratie (MIDEM), Dresden, 2020.
- [8] Statistisches Bundesamt, „www.destatis.de“, 27 05 2026. [Online]. Available: <https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/Archiv/GVAuszugQ/AuszugGV2QAktuell.html>. [Zugriff am 02 06 2026].
- [9] J. MacQueen, „Some methods for classification and analysis of multivariate observations“, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Bd. 1, pp. 281-297, 1967.
- [10] D. C. Montgomery, E. A. Peck und G. G. Vining, INTRODUCTION TO LINEAR REGRESSION ANALYSIS, Hoboken: Wiley, 2012.
- [11] V. Neu, „bpb.de“, Bundeszentrale für politische Bildung, [Online]. Available: <https://www.bpb.de/kurzknapp/lexika/handwoerterbuch-politisches-system/202012/die-linke/>. [Zugriff am 27 06 2026].
- [12] L. Breiman, „Random Forests“, *Machine Learning*, Nr. 45, p. 5–32, 10 2001.
- [13] L. F. Stotzer, C. Erfort, H. Rajski, T. Gschwend, S. Munzert und E. Koch, „An election forecasting model for subnational elections“, *Electoral Studies*, 2025.
- [14] F. Scholles, „Prognose“, *Handwörterbuch der Stadt- und Raumentwicklung*, pp. 1811-1820, 2018.

ERKLÄRUNG

Ich erkläre hiermit, dass ich die vorliegende „Arbeit Demografischer Wandel und Wahlverhalten im Schwalm-Eder-Kreis“ selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

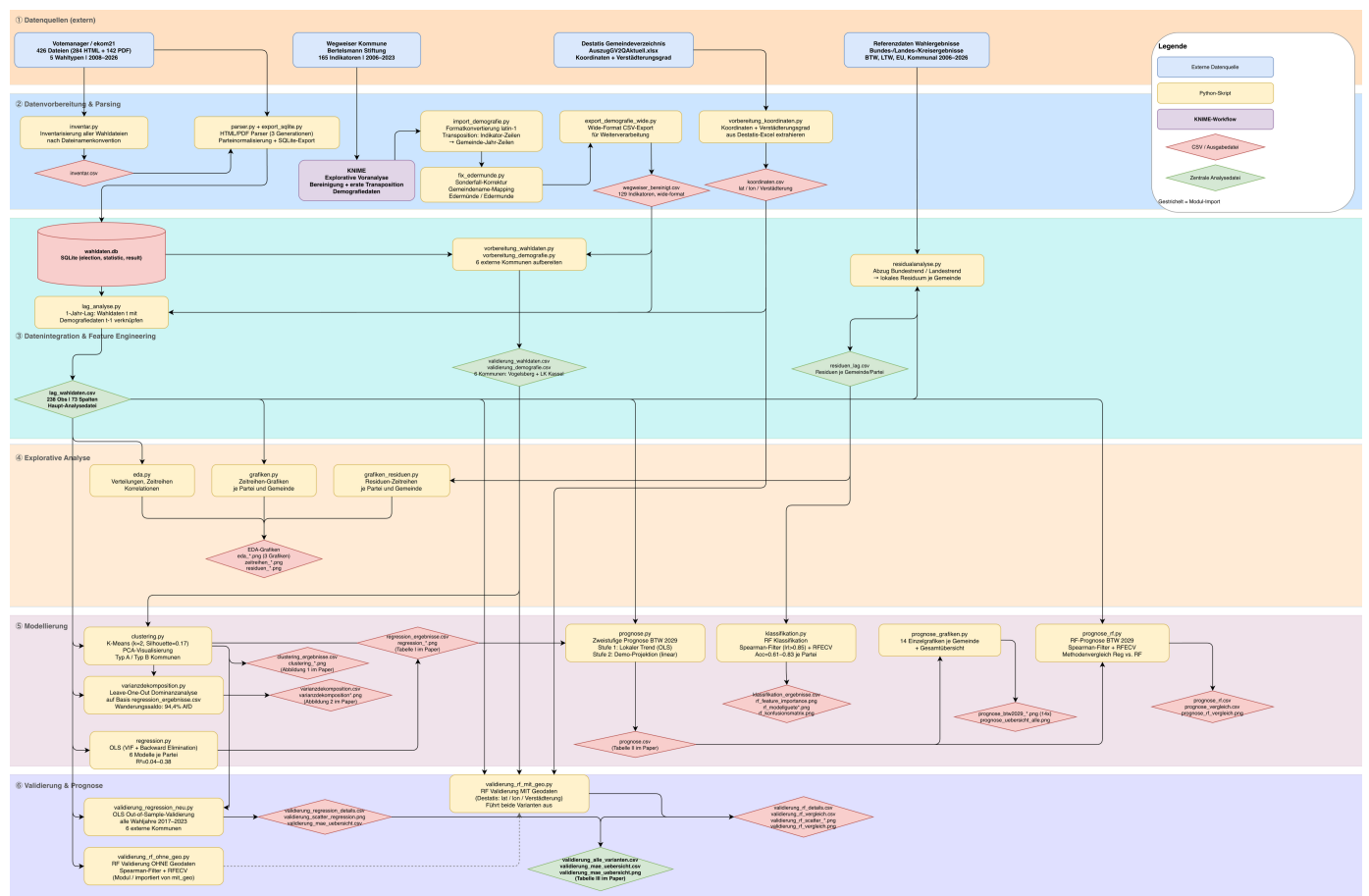
Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Per KI generierte Inhalte habe ich in jedem Fall als solche kenntlich gemacht und im Anhang detailliert beschrieben, wie diese Inhalte erzeugt und auf ihren Wahrheitsgehalt geprüft worden sind.

Die Arbeit wurde nach meiner Kenntnis bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt.

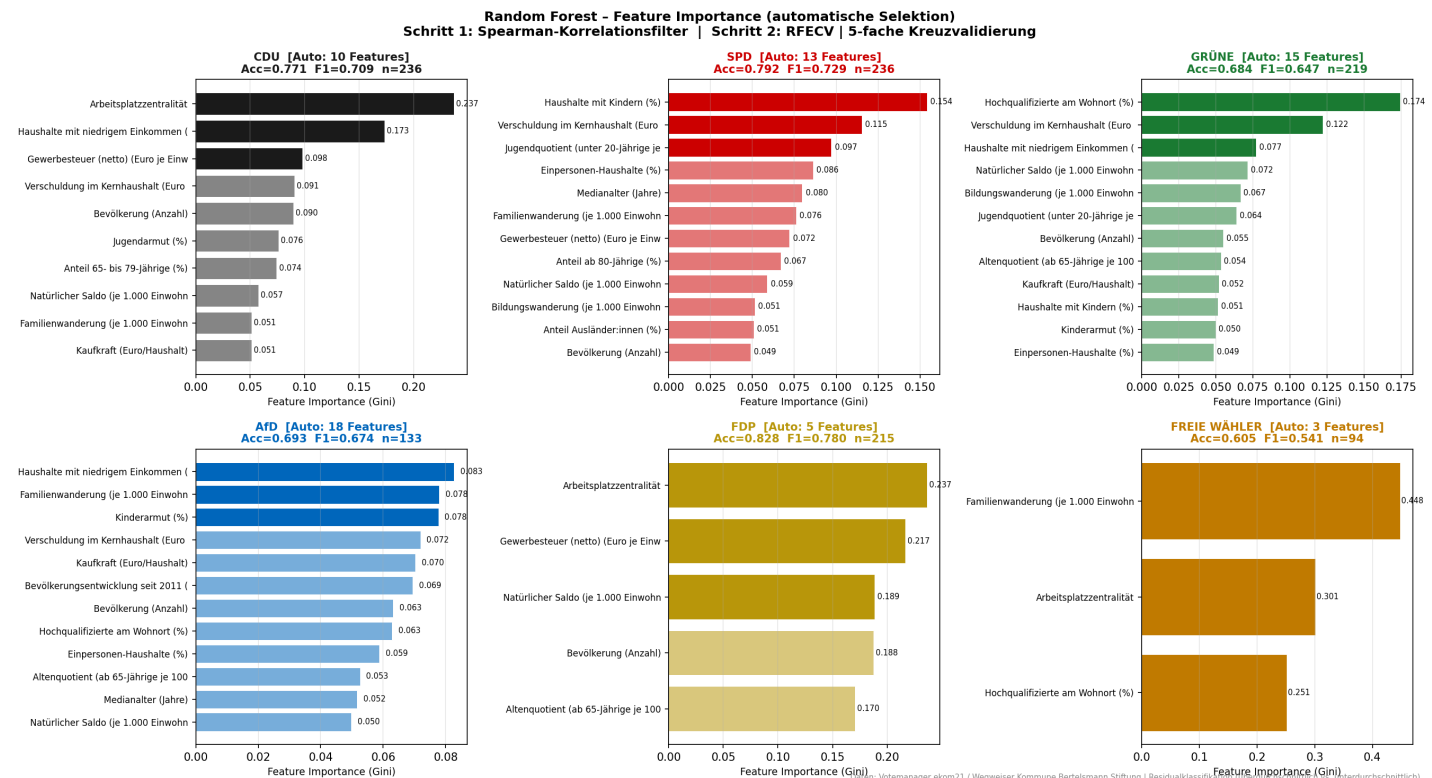
Edermünde, den 01.07.2026 gez. Moritz Schumacher

ANHANG

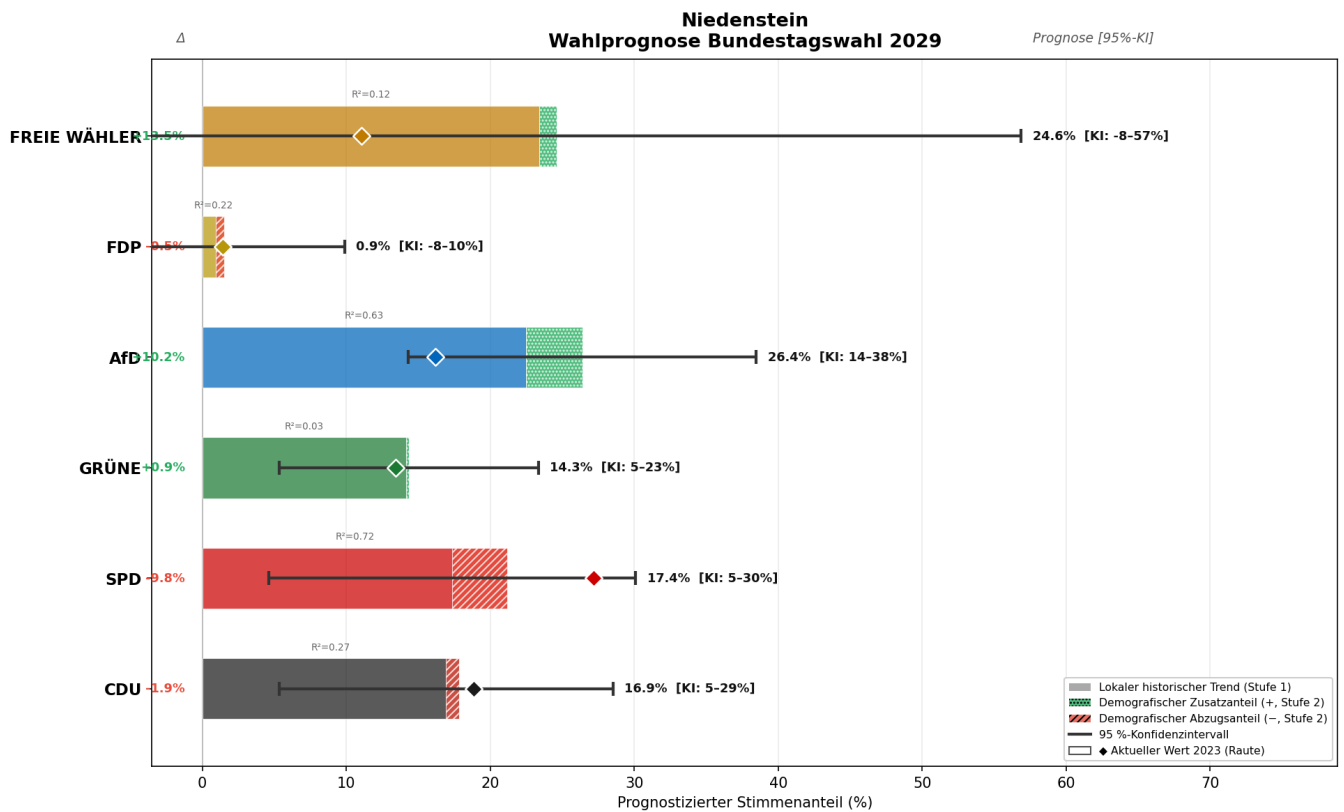
A. Grafiken



Anhang-Abbildung 1 Workflow der Daten-Analyse

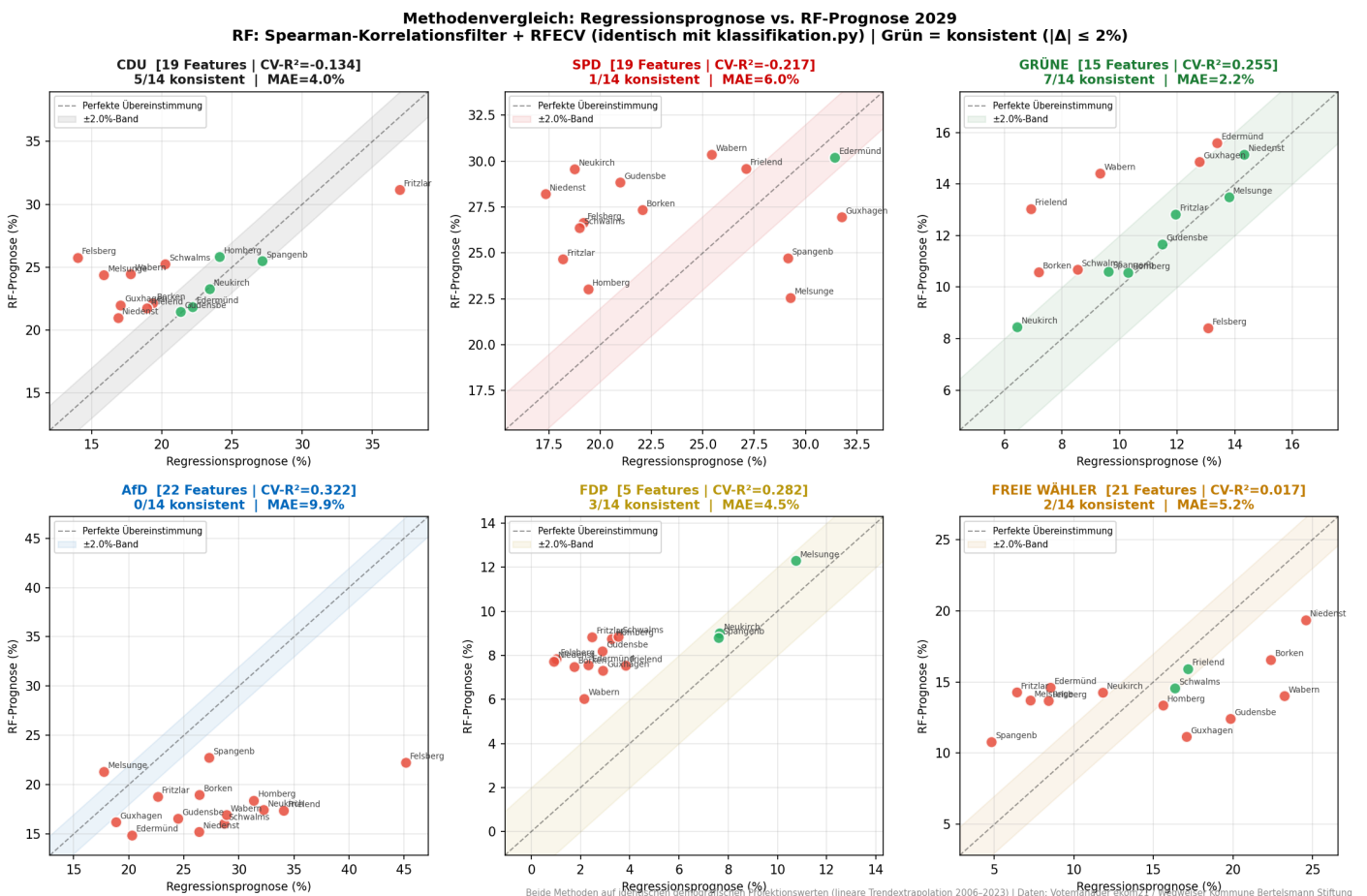


Anhang-Abbildung 2 Feature-Importance des Random Forest für die untersuchten Parteien



Zweistufiges Modell: Lokaler OLS-Trend (hist. Wahlergebnisse) + demografische Trendextrapolation (linear, 2006-2023) | Votemanager ekom21 / Wegweiser Kommune Bertelsmann Stiftung

Anhang-Abbildung 3 Beispielhafte Wahlprognose für die Bundestagswahl 2029 in Niederstein mit sichtbaren demografischen Zusatz- bzw. Abzugsanteilen.



Anhang-Abbildung 4 Methodenvergleich der Regressions- und Random-Forest-Prognose für 2029

B. Beschreibung der KI-Nutzung

Für diese Arbeit wurde mit den Large-Language-Modellen Claude Sonnet 4.6 und Opus 4.7 sowohl der Programmcode für die eingesetzten Skripte (Python), als auch der Workflow erarbeitet. Dabei wurde die grundlegende Aufgabenstellung mit den Modellen erörtert und iterativ ein Vorgehen zum Aufbau einer dem CRISP-DM Prozess entsprechenden Datenverarbeitungsstruktur entwickelt. Für die Quellenrecherche, Validierung von Rechercheergebnissen und Zusammenfassungen wurde weiter Microsoft Copilot eingesetzt. Claude Sonnet wurde zur Übersetzung englischsprachiger Quellen herangezogen. Das unter cdue-dermuende.de abrufbare Dashboard ist mit Claude Code (Opus 4.6) generiert worden. Der Anhang-Abbildung 1 Workflow der Daten-Analyse – wurde mit Hilfe von Claude initial auf Basis des bestehenden Quellcodes erstellt, händisch jedoch nachbearbeitet und qualitätsgesichert.