Datenbasierter Report

Analyse der Entwicklung politischer Diskurse in Deutschland auf Basis von Polibarometer-Daten und eigenen Web-Scraping-Ergebnissen

**1. Einleitung und Forschungsfrage**

In der heutigen politischen Landschaft Deutschlands spielen die Parteien CDU, SPD, AfD und die Grünen eine zentrale Rolle. Ihre mediale Präsenz wird durch eine Vielzahl externer Faktoren beeinflusst, darunter Wahlen, politische Skandale, internationale Krisen oder gesellschaftliche Debatten. Die Art und Intensität dieser Berichterstattung prägen nicht nur das öffentliche Bild der Parteien, sondern korrelieren häufig auch mit deren Umfragewerten. Diese Dynamik zu verstehen, ist von entscheidender Bedeutung, um die Rolle der Medien in politischen Prozessen besser einordnen zu können.

Dieser Report hat zum Ziel, die Entwicklung der medialen Präsenz der genannten Parteien anhand von Wordcounts aus Nachrichtenartikeln zu untersuchen und diese Ergebnisse mit den Umfragedaten des Politbarometers der Forschungsgruppe Wahlen in Beziehung zu setzen.

Die zentrale Forschungsfrage lautet: *Wie beeinflussen politische Ereignisse die mediale Präsenz deutscher Parteien, und welche Auswirkungen hat dies auf deren Umfragewerte?*

Die Analyse beruht auf der Hypothese, dass externe Ereignisse und die mediale Berichterstattung die öffentliche Wahrnehmung und somit auch die politischen Präferenzen der Bevölkerung maßgeblich beeinflussen. Um dies zu untersuchen, wurden große Mengen an Daten verarbeitet, wobei bereits im Vorfeld Schwächen sowohl in der Datengrundlage als auch in der Methodik identifiziert wurden. Diese Schwächen sind essenziell für die Interpretation der Ergebnisse und werden im weiteren Verlauf dieses Reports detailliert erläutert.

**2. Daten und Methoden**

Die Analyse stützt sich auf zwei Hauptdatenquellen. Erstens wurden die Daten des Politbarometers genutzt, das im Zweiwochenrhythmus von der Forschungsgruppe Wahlen erhoben wird. Diese Daten bieten eine langfristige Perspektive auf die Umfragewerte der Parteien. Zweitens wurden Web-Scraping-Daten aus über 50 deutschen Nachrichtenportalen gesammelt, die den zeitlichen Verlauf von Wordcounts und deren Verknüpfung mit politischen Ereignissen dokumentieren.

Die Datenaufbereitung erfolgte in mehreren Schritten. Zunächst wurden die Rohdaten auf fehlende Werte überprüft und bereinigt, was insbesondere bei den Web-Scraping-Daten herausfordernd war, da diese oft inkonsistent formatiert waren. Anschließend wurden die bereinigten Daten in einer SQL-Datenbank zusammengeführt, um eine effiziente Verarbeitung und Analyse zu gewährleisten. Python wurde für die Erstellung von Visualisierungen verwendet, die zeitliche Entwicklungen und Zusammenhänge zwischen den Datenquellen verdeutlichen.

Die Datenbasis weist jedoch Schwächen auf, die die Aussagekraft der Analyse einschränken könnten. Die Politbarometer-Daten basieren auf Stichprobenerhebungen, die nicht immer alle demografischen Gruppen repräsentativ abbilden. Darüber hinaus besteht bei den Web-Scraping-Daten das Risiko, dass relevante Artikel aufgrund ungenauer Keyword-Definitionen oder unterschiedlicher Schreibweisen der Parteinamen nicht erfasst wurden. Dies betrifft insbesondere Abkürzungen oder Synonyme, die in der journalistischen Berichterstattung verwendet werden. Hinzu kommt, dass redaktionelle Haltungen einzelner Medien die Ergebnisse verzerren könnten. Diese methodischen Schwächen müssen bei der Interpretation der Ergebnisse berücksichtigt werden.

**3. Ergebnisse**

Die Analyse zeigt deutliche Unterschiede in der medialen Präsenz der Parteien und deren Entwicklung im betrachteten Zeitraum. Die CDU und SPD zeigen gegenläufige Entwicklungen. Während der Bundestagswahl 2021 sank die mediale Präsenz der CDU infolge negativer Berichterstattung über Armin Laschet, was gleichzeitig mit einem Anstieg bei der SPD korrelierte. Ein prägendes Ereignis war hierbei der sogenannte "Laschet-Lacher" während eines Besuchs im Flutgebiet, der breite mediale Aufmerksamkeit erlangte (FR.de, 2021). Die Grünen erreichten im Mai 2021 einen Höhepunkt in der medialen Aufmerksamkeit, bevor Plagiatsvorwürfe gegen Annalena Baerbock einen starken Rückgang auslösten. Diese Vorwürfe wurden intensiv in den Medien diskutiert und beeinflussten die öffentliche Wahrnehmung der Partei (T-Online.de, 2021). Die AfD hingegen profitierte medial von der Energiekrise und dem sogenannten Heizungsstreit, was auch mit einem Anstieg ihrer Umfragewerte verbunden war. Die Energiekrise führte zu erheblichen Preissteigerungen und Versorgungsunsicherheiten, die in der öffentlichen Debatte stark präsent waren (Statistisches Bundesamt, 2022). Der Heizungsstreit, insbesondere die Diskussion um das Gebäudeenergiegesetz, sorgte für weitere Kontroversen und mediale Aufmerksamkeit (ZDF.de, 2023). Die Visualisierungen zeigen, dass externe Ereignisse wie die Ukraine-Krise oder der Migrationsgipfel nicht nur den Wordcount, sondern auch die Umfragewerte beeinflussten.

Die AfD erzielte während dieser Ereignisse signifikante Zugewinne, während die Grünen in der medialen Präsenz und in den Umfragewerten Verluste verzeichneten. Der Besuch von Friedrich Merz in der Ukraine im Mai 2022 unterstrich die Bedeutung der Ukraine-Krise in der deutschen Politik und Medienlandschaft (CDU.de, 2022). Der Migrationsgipfel und die Veröffentlichung des Migrationsberichts 2022 rückten das Thema Migration erneut in den Fokus der öffentlichen Diskussion (Bundesamt für Migration und Flüchtlinge, 2024).

**4. Diskussion der Ergebnisse und Schwächen**

Die Ergebnisse belegen, dass politische Ereignisse eine starke Wechselwirkung zwischen medialer Präsenz und Umfragewerten der Parteien erzeugen. Dennoch bestehen methodische und datenbezogene Schwächen, die die Aussagekraft der Analyse einschränken.

Eine wesentliche Schwäche liegt in der Qualität der Politbarometer-Daten, die auf Stichprobenerhebungen basieren. Diese sind anfällig für Verzerrungen, insbesondere wenn bestimmte demografische Gruppen nicht ausreichend repräsentiert sind. Die Web-Scraping-Daten weisen ebenfalls Schwächen auf, insbesondere durch die Variabilität der Parteinamen, die zu unvollständigen Datensätzen führen kann. Dies ist besonders problematisch, da politische Ereignisse häufig mit individuellen Akteuren in Verbindung gebracht werden, deren Namen nicht in den Keyword-Sets enthalten sind.

Ein weiterer Schwachpunkt liegt in der potenziellen Verzerrung durch redaktionelle Haltungen der analysierten Medien. Diese könnten dazu führen, dass bestimmte Parteien systematisch über- oder unterrepräsentiert sind. Darüber hinaus gibt es methodische Einschränkungen bei der Verknüpfung von Wordcount-Daten mit den Ergebnissen des Politbarometers. Der direkte kausale Zusammenhang zwischen medialer Präsenz und Umfragewerten ist schwer nachzuweisen und könnte durch eine Vielzahl externer Faktoren beeinflusst werden, die in dieser Analyse nicht berücksichtigt wurden.

**5. Fazit und Handlungsempfehlungen**

Die Analyse zeigt, dass die mediale Präsenz politischer Parteien eng mit aktuellen politischen und gesellschaftlichen Ereignissen verknüpft ist. Die Ergebnisse verdeutlichen, dass externe Ereignisse wie die Ukraine-Krise oder der Heizungsstreit signifikante Auswirkungen auf die öffentliche Wahrnehmung und die Umfragewerte der Parteien hatten.

Um die Aussagekraft zukünftiger Analysen zu verbessern, sollten die Datenmethoden optimiert werden. Eine Erweiterung der Keyword-Datenbank um Synonyme und alternative Schreibweisen könnte helfen, die Vollständigkeit der Datensätze zu gewährleisten. Zudem sollten weitere Datenquellen, beispielsweise Social-Media-Daten, integriert werden, um ein umfassenderes Bild der politischen Dynamik zu erhalten. Schließlich sollten Methoden zur statistischen Gewichtung angewendet werden, um Verzerrungen in den Umfragedaten zu minimieren.