

Agenda-Setting

Eine empirische Analyse der kleinen Anfragen des 18. und 19.
Bundestages

AUTOREN

Marlon SCHUMACHER

Robin GERL

Mail: m.c.schumacher@live.de

Mail: gerl.robin@web.de

Matrikel-Nr.: 2954594

Matrikel-Nr.: 2959667

SEMINAR

INTERESSENGRUPPEN, REGIERUNGEN UND ÖFFENTLICHE MEINUNG

Abgabedatum: 30.03.2019

Inhaltsverzeichnis

1	Einleitung	4
2	Theorie	5
2.1	Issue Ownership	5
2.1.1	Theorie & Hypothesen	5
2.1.2	Aktueller Forschungsstand	7
3	Methode	9
3.1	Datenbeschaffung	9
3.2	Klassifizierung	10
3.2.1	Wortklassifizierung	10
3.2.2	Supervised Machine Learning	11
3.2.3	Sentiment Analyse	14
4	Ergebnisse	15
4.1	Klassifikation	15
4.2	Deskriptive Analyse	19
4.3	Sentiment Analyse	22
5	Fazit	25
6	Literaturverzeichnis	27
7	Appendix	32

Abbildungsverzeichnis

1	Visuelle Darstellung einer linearen Support Vector Machine	13
2	Verteilung der Themen nach Wortklassifikation	15
3	Klassenwahrscheinlichkeiten der Testklassifikation	17
4	Verteilung der Themen nach Klassifikation (RF + SVM)	18
5	Anzahl gestellter Anfragen nach Partei	19
6	Anteil der Anfragen nach Thema und Partei	20
7	Mittelwertvergleich (t-Test) der Sentiments für ausgewählte Themen	23
A1	Genauigkeit in Abhängigkeit von C (SVM)	39
A2	Klassenwahrscheinlichkeiten (vollständige Klassifikation)	39
A3	Worthäufigkeiten der Parteiprogramme	40

Tabellenverzeichnis

1	Richtig kategorisierte Anfragen und Anteil der Klassifikationen	16
A1	CAP Topics und verwendete Wörter für die Kategorisierung	32
A2	Themenspezifischer Fehler nach n_{tree} (Random Forest)	38

1 Einleitung

Die Bundestagswahl 2017 hat das Parlament deutlich verändert. Die AfD ist als drittstärkste Kraft zum ersten Mal in den Bundestag gekommen. Ebenso gelang es der FDP einen Einzug in den Bundestag zu realisieren. Entgegen diesen Erfolgen stehen die Verluste der Union und SPD, welche zusammen über 10% der Stimmen verloren haben (vgl. Tagesschau 2017; Die Welt 2017).

In Hinblick dieser Veränderung und dem Einzug der neuen Parteien mit deren politischen Agenda, ergibt sich die Frage wie sich dies auf die politische Agenda der bisher im Bundestag vertretenen Parteien ausgewirkt hat. Insbesondere der Einzug der AfD ist dabei interessant, da diese Partei, im Gegensatz zur FDP, zum ersten Mal im Bundestag vertreten ist. In dieser Arbeit soll daher versucht werden die folgende Forschungsfrage zu beantworten:

Liegen zwischen den Parteien Unterschiede hinsichtlich ihrer Themenschwerpunkte vor und können durch den Einzug der AfD und FDP Veränderungen festgestellt werden?

Um diese Forschungsfrage zu beantworten müssen einerseits entsprechende Daten herangezogen werden, mit welchen die Identifizierung von Themenschwerpunkten möglich ist. Andererseits müssen klare Themen definiert werden, zwischen welchen differenziert werden soll. Dabei werden die kleinen Anfragen als Datenquelle herangezogen. Grund für die Wahl der kleinen Anfragen ist zum einen die öffentliche Zugänglichkeit, wodurch eine Nachvollziehbarkeit dieser Arbeit gewährleistet ist. Zum anderen ist es auf diese Weise möglich die *Issue Ownerships* auf eine quantitative Weise zu identifizieren. Für die Themenklassifizierung werden die Themen des *Comparative Agendas Project* herangezogen. Da es sich bei den kleinen Anfragen jedoch primär um ein Oppositionsmittel handelt, beschränkt sich die Forschungsfrage auf die Oppositionsparteien des Bundestags.

In Kapitel 2 erfolgt eine theoretische Einbettung sowie die Hypothesengenerierung. Ebenso soll hierbei auf den Forschungsstand der Theorie Issue Ownership eingegangen werden. Die Methodik wird in Kapitel 3 umfänglich beschrieben. Dabei wird die Datenquelle und Datenaufbereitung beschrieben. Die Klassifizierung wird in diesem Kapitel ebenso detailliert beschrieben. Im Anschluss werden in Kapitel 4 die Ergebnisse dargestellt. Dabei wird ebenso die Klassifizierung als solche analysiert. Letztlich erfolgt im Fazit (Kapitel 5) eine kurze Zusammenfassung und kritische Betrachtung der Ergebnisse. Ebenso sollen potentielle Themen für künftige Forschungsarbeiten dargelegt werden.

2 Theorie

2.1 Issue Ownership

In diesem Abschnitt soll zunächst die Theorie der *Issue Ownership* erläutert werden. Dafür werden zunächst die Anfänge erläutert, um dann zu den Kernthesen und Dimensionen der Theorie zu gelangen. Im darauffolgenden Schritt werden die Hypothesen für diese wissenschaftliche Arbeit anhand der *Issue Ownership* Theorie formuliert und erläutert. Anschließend wird der aktuelle Forschungsstand näher beleuchtet. Hierbei soll die Frage beantwortet werden, welche Forschung in der Vergangenheit bereits getätigt wurde.

2.1.1 Theorie & Hypothesen

Die Theorie der *Issue Ownership* erlangte 1983 erstmals in dem Buch *Explaining and predicting elections: Issue effects and party strategies in twenty-three democracies* von Ian Budge und Dennis Farlie in den Politikwissenschaften an Bekanntheit (vgl. Lefevere et al. 2015: 2). Die beiden Autoren sind mit dem Wissenschaftler John R. Petrocik die meistzitiertesten Autoren in der *Issue Ownership* Forschung und können somit als Grundväter dieser Forschung gewertet werden. Ziel dieser Theorie ist es zum einen zu erklären wie sich Parteien und Kandidaten verhalten, die sich in Wahlen auf einzelne Themen fokussieren, welche von den Parteien vorgegeben wurden. Das andere Ziel ist es, das Wahlverhalten bezüglich *Issue Ownership* zu analysieren (vgl. Bélanger & Meguid 2008: 2).

Die Wissenschaftler Budge und Farlie beschreiben *Issue Ownership* als die Wahrnehmung von Wählern, der zufolge eine Partei in einem Themengebiet mehr Kompetenzen besitzt als andere Parteien im jeweiligen Wahlsystem. Dabei spiegelt *Issue Ownership* für Budge und Farlie eine stabile und positive Evaluation von Wählern wieder, nach der Parteien in spezifischen Themenbereichen ihre politischen Ideen umsetzen können (vgl. Stubager 2017: 347). Petrocik beschreibt *Issue Ownership* auf ähnliche Weise. Für ihn entstammt *Issue Ownership* aus der Fähigkeit bestimmter Parteien gewisse Probleme in dem Themenfeld zu lösen. Die Parteien erhalten diese Befähigung, indem sie in diesen Themenfeldern mehr Aufmerksamkeit, Initiative und Innovationen bezüglich der Probleme aufzeigen (vgl. Walgrave et al. 2012: 772). Der erhöhte Ressourcenaufwand führt dazu, dass die Wähler auf lange Sicht diesen Parteien mehr Handlungs- und Wissenskompetenzen in den Themen einräumen. Dabei kann die Affiliation zwischen den Parteien und den Themen aus kurzfristigen und langfristigen Quellen stammen. Petrocik nennt diese Formen *record of incumbent* und *constituencies of the parties* (Petrocik zitiert nach Stubager 2017: 348). Beim *record of incumbent* zählen alle Errungenschaften und Schwierigkeiten von Parteien bezüglich der Lösung von Problemen hinzu. Das Problem hierbei ist, dass unvorhergesehene Ereignisse Schwierigkeiten hervorrufen können, welche nicht von den Parteien ausgelöst wurden und bewältigt werden können, auch wenn die Partei erhöhte Kompetenzen im jeweiligen Themenbereich besitzt. Deswegen wird der *record of incumbent* als

instabil eingeschätzt (vgl. Stubager 2017: 348). Im Kontrast hierzu ist die *constituencies of the parties* sehr viel stabiler. Die Grundlage liegt dabei in den stabilen aber unterschiedlichen sozialen Wählerschaften und der Verbindung zwischen politischen Konflikten und sozialen Strukturen (Petrocik zitiert nach Stubager 2017: 348).

In jüngeren Studien von Walgrave, Lefevere und Tresch aus dem Jahr 2012 wird nochmals zwischen zwei Analysedimensionen der *Issue Ownership* differenziert. Die erste Dimension bezieht sich auf den Kompetenzaspekt von Parteien in den jeweiligen Themen. Die zweite Dimension betrifft den Assoziationsaspekt von Parteien für gewisse Themen (vgl. Walgrave et al. 2012: 772). Beide Dimensionen wurden schon von Budge & Farlie und Petrocik in ihren Werken verwendet. Walgrave et al. kritisieren jedoch, dass beide Dimensionen kombiniert und zeitgleich untersucht werden. Stattdessen sollten beide Dimensionen getrennt gemessen und untersucht werden, um die *Issue Ownership* in vollem Umfang analysieren zu können. Dieses Argument unterstreichen die Autoren mit empirischen Befunden (vgl. Walgrave et al. 2012: 779). Ein anderer relevanter Faktor ist die Frage, ob klare thematische Monopole existieren. Falls eine Partei eine klare Dominanz über ein Thema besitzt, schlägt sich dies auch in Wahlen nieder, wenn das Thema salient ist (vgl. Geys 2012: 4). Zum anderen können auch Parteien ein Thema unter sich aufteilen. Dabei würden verschiedene Wählerbasen mit eigenen Grundansichten verschiedene Parteien bezüglich eines Themas präferieren (vgl. Geys 2012: 5). Jedoch können auch hier Hierarchien bezüglich der Einschätzung der Kompetenz von Parteien und der Wertigkeit der Assoziationen auftreten.

Eines der Hauptprobleme der empirischen Analyse der *Issue Ownership* in Deutschland ist die Messung der Themen. Diese findet in der breiten Forschung durch Survey Fragen statt, mit welchen herausgefunden wird inwiefern Parteien erhöhte Kompetenzen in einem Themenfeld besitzen und ob Parteien mit spezifischen Themen assoziiert werden. Für Deutschland liegen hierfür jedoch keine Daten vor. Mit dieser Problematik setzt sich auch die vorliegende Arbeit auseinander. Können kleine Anfragen für den Bundestag als Messmethode für die *Issue Ownership* Theorie benutzt werden? Ein anderer wichtiger Aspekt, der nicht in der breiten Forschung diskutiert wird, sind die behandelten Themen der Parteien in der politischen Arbeit. Wie positionieren sich die Oppositionsparteien im deutschen Bundestag? Hierbei kann davon ausgegangen werden, dass der Fokus der Parteien sich nicht gleichmäßig auf sämtliche Themen erstreckt. Viel mehr kann es von partiespezifischen Themenschwerpunkten auszugehen.

H1: Die Parteien weisen bei den kleinen Anfragen Themenschwerpunkte vor, welche sich zwischen den Parteien unterscheiden.

Anhand dieser Hypothese soll ermittelt werden, ob sich in der parlamentarischen Arbeit thematischen Präferenzen von Parteien erkennen lassen. Diese Frage könnte in der Parteienforschung interessant werden, um zu ermitteln, inwiefern eine Partei thematische Schwerpunkte im Wahlprogramm setzt, um Wahlen zu gewinnen oder ob die Partei inhärent den Drang besitzt, ein Thema im politischen System zu bedienen. Durch Anfragen im Bundestag können Oppositionsparteien stetig Informationen aus den Regierungsfractionen erhalten, welche letztendlich in kommenden Wahlkämpfen instrumentalisierbar sind.

Das Instrument der kleinen Anfragen wird im Bundestag hauptsächlich nur von Oppositionsparteien genutzt. Im 18. Bundestag befanden sich nur die Parteien Bündnis 90/Die Grünen und Die Linke in der Opposition. In der Regierung waren die Parteien SPD und CDU/CSU vertreten. Im 19. Bundestag blieb die Regierungskonstellation gleich, wohingegen zu den bestehenden Oppositionsparteien die Parteien AfD und FDP hinzu kamen. Nach der *Issue Ownership* sollten mit den neuen Parteien ebenso neue Themen im 19. Bundestag vertreten sein. Ausgehend dessen und der potentiellen Dominanz beider Parteien in spezifischen Themengebieten, kann vermutet werden, dass die bisher vertretenen Oppositionsparteien sich auf diese neu hinzugekommenen Themen verstärkt fokussieren.

H2: Die Themenschwerpunkte der Parteien Die Linke und Bündnis 90/Die Grünen haben sich im 19. Bundestag in Richtung der dominanten Themen der FDP und AfD verschoben.

Eine letzte zu untersuchende Frage besteht darin, ob sich die Ansichten von Parteien aus unterschiedlichen politischen Richtungen bezüglich der selben Themen unterscheiden. Ausgehend von Geys (2012) können mehrere Parteien ein Thema dominieren, jedoch können dabei unterschiedliche Ansichten vertreten sein. Bei einer deutlichen Differenz sollte dies durch das Sentiment beider Parteien deutlich werden.

H3: Parteien weisen bei den kleinen Anfragen zu gleichen Themen unterschiedliche Sentiments auf.

2.1.2 Aktueller Forschungsstand

Die Theorie der *Issue Ownership* ist eine relativ neue Herangehensweise, um das Wahlverhalten in politischen Systemen zu analysieren. Hierbei wird relativ schnell erkennbar, dass die Forschung aus theoretischer Sicht stagniert. In den meisten Studien werden nur empirische Überprüfungen getätigt. Eine tiefergehende Betrachtung der Theorie und Modifizierungen dieser findet kaum statt. Die Theorie von Budge & Farlie und Petrocik bleibt dementsprechend die Grundlage für alle zukünftigen Forschungen. Nur die Autoren Walgrave et al. versuchten sich an einer Modifizierung der Theorie und führten erste empirische Überprüfungen durch (Walgrave et al. 2012). Benny Geys entwickelte ebenfalls einen theoretischen Ansatz, versuchte jedoch nicht seine Annahmen in einer empirischen Studie zu überprüfen (Geys 2012).

In der empirischen Anwendung wurden im Gegenzug bereits diverse Studien durchgeführt, um die Theorie der *Issue Ownership* zu überprüfen. Hierbei haben Forscher verschiedene Elemente und Schnittstellen der Theorie überprüft. In einer früheren Studie versuchten beispielsweise die Autoren Ansolabehere und Iyengar herauszufinden, ob für Wahlkampfserfolge die *Issue Ownership* Theorie oder die *Riding the wave* Theorie bessere Erklärungen erzielen (Ansolabehere & Iyengar 1994). Kleinnijenhuis stellte ebenfalls einen Theorievergleich mit den Theorien *Issue Ownership*, *Issue convergence* und *Spatial Proximity* an (Kleinnijenhuis & Lefevere). Im Gegenzug stellte sich Van Camp die Frage, inwiefern *Issue Ownership* eine Determinante für Medienberichterstattung ist und Walgrave und De Swert untersuchten, ob *Issue Ownership* von den Parteien oder den Medien stammt (Van Camp 2018; Walgrave & De Swert 2007). Gunther et al. versuchten zu evaluieren, wie stark BewerberInnen und deren Themen überhaupt in den Medien sichtbar sind (Günther et al. 2017). Ein großer Teil der Forschung bezieht sich jedoch letztlich auf Parteien und wie *Issue Ownership* sich auf Wahlkampfstrategien, Parteienwettbewerb, Wahlkampfthemen und Parteipräferenzen auswirkt (Holian 2004; Green & Hobolt 2008; Green-Pedersen 2007; Stubager & Slothuus 2013; Stubager & Seeberg 2016; Walgrave et al. 2014; Petrocik et al. 2003). Die meiste Forschung wurde jedoch auf der Ebene der Wähler getätigt. Hierbei sollte untersucht werden, wie Faktoren wie Wissen, Wahldebatten, Themensalienz und Wahrnehmung einen Einfluss auf das Wahlverhalten oder auf Faktoren der *Issue Ownership* besitzen und wie dies gemessen werden kann (Bélanger & Meguid 2008; Dejaeghere & Van Erkel 2017; Kaufmann 2004; Neundorff & Adams 2018; Stubager 2017; Van der Brug 2004).

Äußerst aufschlussreiche Studien wurden von Bélanger und Christensen et al. veröffentlicht. Bélanger untersuchte die Dynamiken innerhalb des *Issue Ownership* Konzepts und Christensen et al. untersuchten Änderungen von *Issue Ownership* innerhalb eines politischen Systems (Bélanger 2003; Christensen et al. 2015). Alle Studien behandeln jedoch vordergründig westliche politische Systeme. Hierunter fallen die Vereinigten Staaten von Amerika, sowie Kanada, Schweden, Niederlande, Großbritannien, Deutschland, Dänemark und Belgien.

3 Methode

Im Folgenden soll die methodische Vorgehensweise erläutert werden, welche die Datenbeschaffung, die Klassifizierung sowie die anschließende Analyse beschreiben soll. Da es keinen Datensatz gibt, welcher sämtliche kleine Anfragen sowie deren thematische Klassifizierung enthält, mussten die Daten eigenständig erhoben und klassifiziert werden. Bei den kleinen Anfragen handelt es sich um Drucksachen des deutschen Bundestags, welche durch das *Dokumentations- und Informationssystem für Parlamentarische Vorgänge* öffentlich zugänglich sind (vgl. DIP 2019). Diese Datenbank wird folglich als Datenquelle fungieren. Für sämtliche nachfolgend beschriebenen Schritte wurde die Programmiersprache R genutzt.¹

3.1 Datenbeschaffung

Ausgehend der angeführten Punkte wurden sämtliche Drucksachen des 18. und 19. Bundestages automatisiert heruntergeladen.² Da der 19. Bundestag nach wie vor besteht, musste ein Datum festgelegt werden, ab welchen keine weiteren Drucksachen mehr berücksichtigt werden. Der Zeitraum des 18. Bundestag beläuft sich vom 22. Oktober 2013 bis zum 24. Oktober 2017 und für den 19. Bundestag wurde der Zeitraum vom 24. Oktober bis zum 31. Dezember 2018 festgelegt. Für den 18. Bundestag wurden demnach 13.705 Drucksachen und für den 19. Bundestag 6.896 Drucksachen heruntergeladen. Zusätzlich wurde die Trefferliste als CSV heruntergeladen, welche die Nummer der Drucksache sowie Titel und Datum der Drucksache enthält. Da jede Drucksache als PDF gespeichert wurde und ein standardisiertes Format aufweist, konnten die interessierten Bestandteile des Textes extrahiert werden.³

Zunächst wurde überprüft, ob es sich bei den Drucksachen um eine Anfrage handelt. Sofern es sich um keine Anfrage handelte, wurde die entsprechende Drucksache gelöscht. Anschließend wurden zudem die großen Anfragen ebenso gelöscht, da sich die Gesamtanzahl dieser über den gesamten Zeitraum auf unter 30 beläuft.⁴ Im 18. Bundestag wurden 3.951 kleine Anfragen gestellt und im 19. Bundestag 2.398. Der Datensatz beinhaltet somit 6349 kleine Anfragen. Anschließend wurde die Partei für jede Anfrage ermittelt, welche stets am Ende des Titels steht. Zudem wurde der Inhalt jeder Anfrage extrahiert, welcher für die spätere Kategorisierung herangezogen wird. Der hierdurch erstellte Datensatz wurde durch die heruntergeladen CSV-Datei mit Titel und Datum ergänzt.

¹Der Code für alle Auswertungen der Hausarbeit ist auf GitHub unter https://github.com/MCStatistic/SeminarPaper_ASGB einsehbar.

²Hierfür wurde eine Funktion in R geschrieben, mit welcher jede Drucksache als PDF vom offiziellen Server heruntergeladen wurde. Die Dateien wurden dabei unter der Nummer der Drucksache sowie des jeweiligen Bundestags gespeichert um so eine klare Zuordnung gewährleisten zu können.

³Hierfür wurde das Package *pdftools* (vgl. Ooms 2019) genutzt.

⁴Da aus diesem Grund die großen Anfragen in dieser Arbeit nicht berücksichtigt werden, wird nachfolgend zwecks der Lesbarkeit von *Anfragen* geschrieben. Dieser Begriff wird in dieser Arbeit äquivalent zu *kleinen Anfragen* genutzt.

3.2 Klassifizierung

Im Hinblick auf die Forschungsfrage müssen die Anfragen nach Themen klassifiziert werden. Dabei müssen die Themen zuvor klar definiert werden, welche für die Klassifizierung herangezogen werden sollen. Hierbei soll sich verstärkt an der Arbeit von Baumgartner, Jones und Wilkerson orientiert werden, welche für die USA und der dortigen nationalen Politik ein Kodiersystem erstellt haben (vgl. Baumgartner et al. 2006: 970). Neben dem Kodiersystem, welches für die US-Politik erstellt wurde, existiert eine neuere Version, welche allgemeine Kodierrichtlinien beinhaltet (vgl. Comparative Agendas Project 2019). Einerseits kann das Kodiersystem leicht auf verschiedene Länder angewendet werden. Andererseits ergibt sich durch die Verwendung dieses Kodiersystems die Möglichkeit der Vergleichbarkeit mit anderen Ländern, was jedoch in dieser Arbeit nicht weiter von Interesse ist (vgl. John 2006: 983).

Anhand dieses Kodierungssystems werden die Anfragen in dieser Arbeit klassifiziert. Dabei enthält das Kodiersystem 21 Hauptthemen und 220 Subthemen. Zwischen den Subthemen wird bei der nachfolgenden Klassifizierung jedoch nicht differenziert. Diese Entscheidung ergibt sich aus zwei Gründen. Zum einen stellen die Hauptthemen bereits eine ausreichende Differenzierung zur Beantwortung der Forschungsfrage dar, wodurch eine zusätzliche Differenzierung zwischen den Subthemen keinen Mehrwert bringen würde. Zum anderen ist eine derart detaillierte Klassifizierung hinsichtlich der methodischen Vorgehensweise schwer zu realisieren, da es für einige der 220 Subthemen zu wenig Anfragen gibt. Eine Übersetzung der Kodierung ist im Anhang in Tabelle A1 enthalten. Die Tabelle enthält die Hauptthemen sowie deren Subthemen. Ebenso sind in der Tabelle die Wörter enthalten, welche für den ersten Schritt der Klassifizierung herangezogen wurden.

Die Klassifizierung der Anfragen erfolgt in zwei Schritten. Im ersten Schritt werden diese anhand von Wörtern klassifiziert und anschließend erfolgt eine Validierung dieser Klassifizierung. Im zweiten Schritt erfolgt eine Klassifizierung der Anfragen, die nicht durch den ersten Schritt klassifiziert werden konnten. Hierfür werden zwei Verfahren des Supervised Machine Learnings verwendet. Auf diese beiden Schritte wird im Folgenden detailliert eingegangen.

3.2.1 Wortklassifizierung

Anhand der Subthemen wurden für jedes Hauptthema Wörter identifiziert, welche ausschließlich auf das jeweilige Hauptthema zutreffen und sich thematisch von den übrigen Hauptthemen⁵ klar abgrenzen sollten. Dabei wurden die Wörter unter anderem auch durch das Querlesen einiger Anfragen ermittelt. Um mit Hilfe der Wörter eine zuverlässige Klassifizierung zu ermöglichen, wurde bei diesem Schritt lediglich der Titel der jeweiligen Anfrage herangezogen. Einerseits fungiert der Titel als eine Beschreibung des Inhalts, wodurch dieser entsprechend klar und prägnant formuliert sein muss. Andererseits ist davon auszugehen, dass im Inhalt der Anfrage

⁵Da keine Differenzierung zwischen den Subthemen erfolgt und nachfolgend nur noch eine Differenzierung zwischen den Hauptthemen relevant ist, wird ab sofort nur noch von Themen geschrieben. Die Begriffe *Hauptthema* und *Thema* werden in dieser Arbeit daher äquivalent genutzt.

ein Bezug zu anderen Themen vorkommen kann. Eine derartige Klassifizierung anhand der Inhalte würde folglich zu einem tendenziell unzuverlässigem Ergebnis führen.

Da im zweiten Schritt die bereits klassifizierten Anfragen als Grundlage dienen, wurde die Zuverlässigkeit der soeben beschriebene Klassifizierung überprüft. Dabei wird zunächst analysiert, ob Anfragen mittels der verwendeten Klassifizierung zu mehreren Themen zugeordnet werden könnten. Anschließend müssen diese identifizierten Anfragen händisch klassifiziert werden. Zusätzlich wurden die Anfragen, welche nur einem Thema zugeordnet wurden, ebenso überprüft. Auf diese Weise soll eine hohe Zuverlässigkeit der Klassifizierung gewährleistet sein.

3.2.2 Supervised Machine Learning

Durch die Klassifizierung mittels themenspezifischer Wörter können nicht alle Anfragen klassifiziert werden, wodurch ein weiterer Schritt notwendig sein wird. Ausgehend der klassifizierten Anfragen ergibt sich die Möglichkeit Methoden des Supervised Machine Learnings heranzuziehen. Dabei sollen zwei Klassifikationsverfahren verwendet werden. Es handelt sich um den Klassifikator *Random Forest* und um das Klassifikationsverfahren mittels *Support Vector Machines*. Die grundlegenden Konzepte beider Verfahren sollen dabei zunächst beschrieben werden.

Ein Random Forest wird nach Breiman (2001) wie folgt definiert:

Ein Random Forest ist ein Klassifikator, welcher aus einer Menge von *Tree*-strukturierten Klassifikatoren besteht $\{h(\mathbf{x}, \Theta_k), k = 1, \dots\}$. Dabei handelt es sich bei $\{\Theta_k\}$ um unabhängige zufällige Vektoren, welche eine identische Verteilung aufweisen. Jeder *Tree* gibt dabei eine Stimme für die beliebteste Klasse für \mathbf{x} ab (vgl. Breiman 2001: 6).

Der Algorithmus, welcher beim Random Forest⁶ zum Einsatz kommt, kann dabei wie folgt beschrieben werden (vgl. Liaw et al. 2002: 18; vgl. Friedman et al. 2017: 588):

1. Aus dem Datensatz werden n_{tree} Samples mittels Bootstrap gezogen.
2. Für jedes dieser Bootstrap-Samples wird ein *unpruned Tree* erstellt.⁷ Für die Erstellung der *Trees* wird der Gini Index G verwendet, wobei T das Training Set, C_i die Klassen und $f(C_i, T)/|T|$ die Wahrscheinlichkeit, dass ein Fall zu der Klasse C_i gehört, darstellt (vgl. Pal 2005: 218; vgl. Chan & Paelinckx 2008: 3002):

$$G = \sum_{j \neq i} \sum (f(C_i, T)/|T|)(f(C_j, T)/|T|)$$

⁶Grundlegend handelt es sich um eine Ansammlung von Decision Trees, welche hier jedoch nicht im Detail ausgeführt werden sollen. Dabei sei auf die Arbeit von Breiman et al. (1984) und Banfield et al. (2007) verwiesen.

⁷Um ein Overfitting bei Decision Trees zu vermeiden, wird die Methodik des *Prunings* verwendet (siehe Bradford et al. 1998; Mingers 1989; Mehta et al. 1995). Dies ist beim Random Forest jedoch nicht erforderlich, da durch das Bootstrapping (oder auch *Bagging*) der Fehler bereits signifikant reduziert wird (siehe hierzu Breiman 1996; Prasad et al. 2006: 184). Die Gefahr eines Overfittings ist bei einem Random Forest daher kein Problem (vgl. Breiman 2001).

3. Abschließend können mit einem solchen trainiertem Modell neue Daten mittels Aggregation den Vorhersagen der einzelnen n_{tree} *Trees* klassifiziert werden. Wenn $\hat{C}_b(x)$ die Klassenvorhersage des b ten Random Forest Tree ist, so ist für $\hat{C}_{rf}^B(x)$ = Mehrheit der Stimmen $\{\hat{C}_b(x)\}_1^B$ ($B = b_1, \dots, b_n$).

Der Vorteil von Random Forest gegenüber einzelnen Decision Trees ist eine signifikant bessere Performance. Zudem ist es weniger anfällig hinsichtlich *Rauschen* (vgl. Ali et al. 2012: 274). Jedoch kann bei (stark) ungleicher Verteilung der Klassen ein Bias in Richtung der *Majority Class* vorliegen (vgl. Boulesteix et al. 2012: 18 f.).⁸

Neben Random Forest gibt es u.a. die Support Vector Machines. Da diese eine etablierte und verbreitete Methode der Klassifikation darstellen und zudem als äußerst genau gelten (vgl. Ben-Hur & Weston 2010: 223), soll in dieser Arbeit ebenso diese Methode herangezogen werden. Ursprünglich wurden Support Vector Machines für die binäre Klassifikation entwickelt (vgl. Cortes & Vapnik 1995: 290). Prinzipiell können Support Vector Machines nur binäre Klassifikationsprobleme lösen, jedoch gibt es Möglichkeiten mit welchen auch eine Multiklassifikation umgesetzt werden kann.

Das grundlegende Konzept von Support Vector Machines soll anhand der Arbeit von Cortes & Vapnik (1995) in Kürze dargelegt werden. Dabei definiert sich eine Support Vector Machine über drei Eigenschaften:

1. Es erfolgt eine Klassenseparierung durch das ermitteln einer *optimalen Hyperebene*. Bei einer binären Klassifikation und einem Input-Vektor von $x = (x_1, \dots, x_n)$ erfolgt die Zuweisung von x zur positiven Klasse, wenn $f(x) \geq 0$ und zur negativen Klasse, wenn $f(x) < 0$. Dabei ist $f(x) = \sum_{i=1}^n w_i x_i + b$, wobei w der gewichtete Vektor und b den Bias darstellt. Die Hyperebene definiert sich letztlich durch $\sum_{i=1}^n w_i x_i + b = 0$ (vgl. Cristianini & Shawe-Taylor 2000: 9 f.). Die *optimale Hyperebene* separiert dabei die Daten mit der größtmöglichen bzw. der *optimalen Marge* zwischen den Klassen (siehe Abbildung 1). Fälle, welche auf den Grenzen der Marge liegen, stellen die *support vectors* dar (vgl. Cortes & Vapnik 1995: 275 ff.).
2. Ein aus Punkt 1 resultierendes Problem ist, dass das Training Set nicht immer ohne Fehlklassifizierungen separiert werden kann. Bei einer harten Marge könnte auf diese Weise keine Modellierung erfolgen. Um eine Separation mit möglichst wenig Fehlern von ξ_i zu ermöglichen, wird bei der C-Support Vector Klassifikation⁹ die Konstante C verwendet. Auf diese Weise wird eine *soft margin Hyperplane* erstellt, wobei mit größerem C die Marge kleiner wird. (vgl. Cortes & Vapnik 1995: 280 ff.; Chang & Lin 2011: 3):

⁸Zwar gibt es Möglichkeiten um diesem Problem entgegenzuwirken, jedoch soll stattdessen eine Kombination der beiden herangezogenen Klassifikationsverfahren verwendet werden, sofern dies erforderlich sein sollte.

⁹C-Support Vector Klassifikatoren werden auch in dieser Arbeit herangezogen. Aus diesem Grund werden andere Lösungen für das angeführte Problem nicht thematisiert.

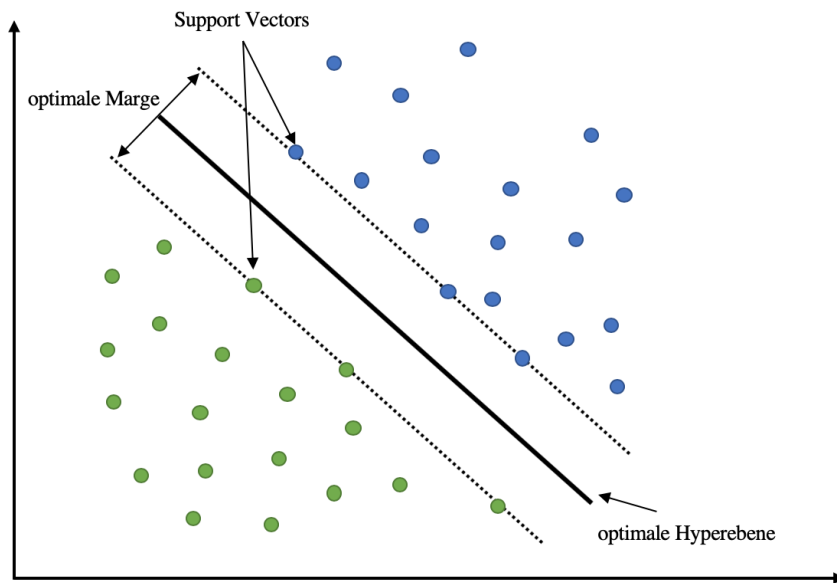
$$\min_{w,b,\xi} \quad \frac{1}{2}w^2 + CF \sum_{i=1}^l \xi_i^\sigma$$

bezogen auf $y_i(w * x_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l,$

$$\xi_i \geq 0, \quad i = 1, \dots, l.$$

3. Sofern eine lineare Separation der Daten im *Input Space* nicht möglich ist, erfolgt eine Ermittlung der Hyperebene auf einem höher dimensional *Feature Space*. Auf diese Weise werden die Daten auf einen anderen Raum projiziert, wodurch eine lineare Separierung ermöglicht wird. Dies wird durch Kernel-Techniken¹⁰ realisiert (vgl. Cortes & Vapnik 1995: 282 ff.; Cristianini & Shawe-Taylor 2000: 26 ff.).

Abbildung 1: Visuelle Darstellung einer linearen Support Vector Machine



Anmerkung: Darstellung anlehnend an Cortes & Vapnik 1995 (S. 275)

Nachdem das grundlegende Konzept von Support Vector Machines aufgezeigt wurde, wird deutlich weshalb prinzipiell nur eine binäre Klassifikation möglich ist. Für eine Multiklassifikation wurden jedoch Methoden entwickelt, wodurch Support Vector Machines ebenso für solche Klassifikationsprobleme verwendet werden können. In dieser Arbeit soll die Methode *one-against-one* genutzt werden. Hierbei werden für k -Klassen $k(k - 1)/2$ Klassifikatoren bzw. Support Vector Machines erstellt. Ebenso wie beim Random Forest erfolgt die finale Zuweisung der Klasse mittels Voting, wobei x zu der Klasse zugewiesen wird, für welche die größte Anzahl an Stimmen vorliegt (vgl. Chang & Lin 2011: 29 f.). Darüber hinaus ist es möglich ebenso die Klassenwahrscheinlichkeit zu berechnen (siehe Wu et al. 2004).

Die soeben beschriebenen Algorithmen können in R mit dem Package `RTextTools` (Jurka et al. 2015; Collingwood et al. 2013) angewendet werden. Hierbei handelt es sich um ein Package,

¹⁰Hierbei soll vorerst nicht näher eingegangen werden, da es viele verschiedene Kernel-Techniken gibt. Für eine allgemeine detaillierte Übersicht diverser Kernel-Techniken bietet sich die Arbeit von Cristianini & Shawe-Taylor (2000) an. Für diese Arbeit wird ein lineares Kernel verwendet.

welches explizit für die Klassifikation von Texten entwickelt wurde und verschiedene Packages vereint. So ist für Random Forest das Package `randomForest` (Breiman & Cutler 2018) und für Support Vector Machines das Package `e1071` (Meyer et al. 2019) implementiert.¹¹ Für die Klassifikation werden die bereits klassifizierten Anfragen als Input herangezogen. Zugleich werden die Modelle mittels der Inhalte der Anfragen trainiert. Dabei müssen die Texte der Inhalte zuvor von den Stoppwörtern, den Punctuationen und den Wörtern, welche nur bei äußerst wenigen Dokumenten vorkommen, bereinigt werden, da dies für den Inhalt und der Klassifikation nicht relevant ist. Zugleich müssen die Texte der Inhalte in ein strukturiertes Format umgewandelt werden, was durch die Erstellung einer Document-Term-Matrix erfolgt (vgl. Meyer et al. 2008: 5 & 23). Vor der richtigen Klassifikation soll anhand der bereits klassifizierten Anfragen die Performance beider Klassifikationsverfahren verglichen werden.

Im Anschluss an die Klassifizierung kann untersucht werden in welcher Weise sich die Parteien hinsichtlich der thematischen Schwerpunkte voneinander unterscheiden. Ebenso können die Unterschiede zwischen dem 18. und 19. Bundestag herausgearbeitet werden.

3.2.3 Sentiment Analyse

Da davon auszugehen ist, dass die Parteien hinsichtlich der Themen unterschiedliche Sentiments aufweisen können, soll ebenso eine Sentiment Analyse durchgeführt werden. Mittels eines Wörterbuchs soll das durchschnittliche *Sentiment* der einzelnen Anfragen ermittelt werden. Hierbei wird das öffentlich zugängliche deutsche Wörterbuch *SentiWS* verwendet (SentiWS 2019). Dieses enthält Wörter mit positiven und negativen (numerischen) Konnotationen. Eine detaillierte Beschreibung dieses Wörterbuchs sowie der verwendeten Methoden zwecks Erstellung ist in Remus et al. (2010) und Goldhahn et al. (2012) vorzufinden. Für die Sentiment Analyse wird das Package `SentimentAnalysis` genutzt (Feuerriegel & Proellocks 2019). Hiermit wird für jede Anfrage ein Sentiment ermittelt:¹²

$$Sentiment = \frac{\sum_{i=1}^n p_i - \sum_{i=1}^n n_i}{\sum_{i=1}^n w_i}$$

Hierbei wird die Differenz zwischen der Summe positiver, p , und negativer Wörter, n , durch die Gesamtanzahl der Wörter, w , dividiert. Auf diese Weise wird ebenso die Größe des Dokuments berücksichtigt. Aufgrund der hohen Anzahl der Themen soll die Analyse dabei auf die am häufigsten auftretenden Themen beschränkt werden.

¹¹Die detaillierten Algorithmen, welche bei den Packages zum Einsatz kommen, können für `randomForest` in Breiman (2001) und für Support Vector Machines in Chang & Lin (2011) vorgefunden werden.

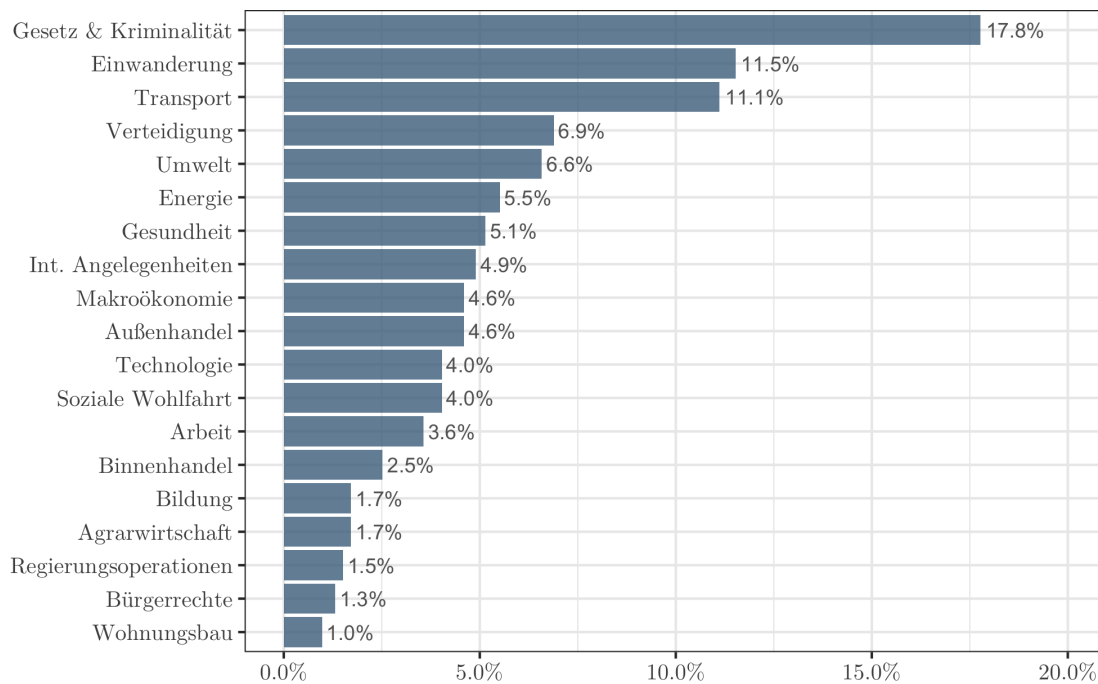
¹²Die hier aufgeführte Berechnung entspricht der Logik, welche im Package `SentimentAnalysis` unter der Funktion `ruleSentiment()` vorzufinden ist. Dies kann im Reference Manual des Packages eingesehen werden (Feuerriegel & Proellocks 2019).

4 Ergebnisse

4.1 Klassifikation

Durch die Wortklassifizierung konnten 2726 Anfragen eindeutig klassifiziert werden. Weitere 253 Anfragen konnten durch die Wortklassifizierung zu 2 oder 3 Themen zugeordnet werden. Diese Anfragen wurden manuell richtig klassifiziert. Darüber hinaus wurden die Anfragen, welche durch die Wortklassifizierung nur einem Thema zugeordnet werden konnten, überprüft und bei wenigen Fällen korrigiert. Anschließend wurden weitere wenige Anfragen manuell kategorisiert um die Datengrundlage für die nachfolgenden Schritte der Kategorisierung weiter vergrößern zu können. Auf diese Weise wurden 3047 der insgesamt 6349 Anfragen kategorisiert. In Abbildung 2 ist die Verteilung der Themen ersichtlich.

Abbildung 2: Verteilung der Themen nach Wortklassifikation



Dabei wird deutlich, dass die Themen *Gesetz & Kriminalität*, *Einwanderung* und *Transport* jeweils einen Anteil von mindestens 10% der klassifizierten Themen aufweisen. Gleichzeitig liegen Themen wie *Wohnungsbau*, *Bürgerrechte* und *Regierungsoperationen* vor, welche jeweils nur einen Anteil von 1,0% bis 1,5% ausmachen. Diese Form der ungleichen Verteilung der Themenanteile kann bei den nachfolgenden Klassifikationsschritten unter Umständen dafür sorgen, dass die Themen unterschiedlich gut klassifiziert werden können.

Bevor eine Klassifikation anhand aller bereits klassifizierten Anfragen umgesetzt wird, soll zunächst die Performance beider Klassifikationsverfahren untersucht werden. Einerseits können auf diese Weise wichtige Parameter der jeweiligen Modelle optimiert werden. Andererseits können so spezifische Grenzwerte festgelegt werden, die für eine zuverlässige Klassifizierung notwendig sind.

Hierfür wurden mittels der 3047 klassifizierten Anfragen ein Testdatensatz erstellt, wobei 2300 dieser Anfragen für das Training der Modelle verwendet wurden und die übrigen 747 Anfragen zur Überprüfung genutzt wurden. Diese Einteilung erfolgte mit vorheriger Randomisierung des erstellten Testdatensatzes.

Da für Random Forest (RF) und Support Vector Machines (SVM) jeweils Klassenwahrscheinlichkeiten berechnet werden, ist es möglich diese als Grenzwert heranziehen zu können. Bei einem hohen Grenzwert steigt zwar die Zuverlässigkeit der Klassifikation, jedoch sinkt ebenso der gesamte Anteil an klassifizierten Anfragen. In Tabelle 1 ist für jedes Modell der Anteil der richtig klassifizierten Anfragen sowie der Gesamtanteil der klassifizierten Anfragen ersichtlich. So zeigen sich keine bedeutenden Unterschiede hinsichtlich der genutzten Anzahl an *Trees* bei RF. Für SVM wurden unterschiedliche C-Parameter getestet, wobei ein Wert von 10 die beste Performance erzielt hat (vgl. Abbildung A1).

Tabelle 1: Richtig kategorisierte Anfragen und Anteil der Klassifikationen

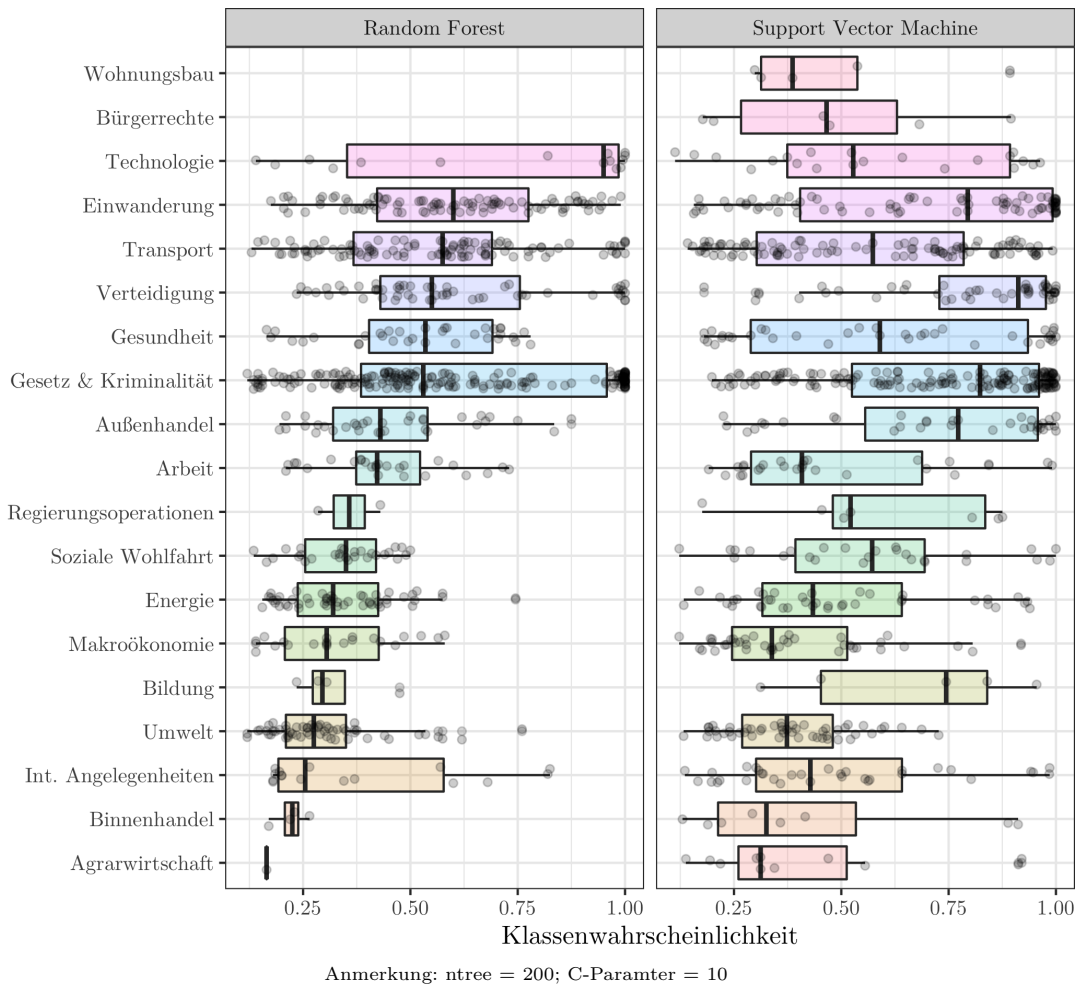
p	RF_{200}		RF_{500}		RF_{1500}		$SVM_{C=10}$	
	Richtig	Anteil	Richtig	Anteil	Richtig	Anteil	Richtig	Anteil
0.3	83,89	74,83	84,96	73,89	85,08	72,69	80,91	76,45
0.4	90,07	58,00	89,47	58,50	90,42	57,29	85,54	64,79
0.5	94,21	41,63	93,08	42,57	93,13	42,84	88,30	52,61
0.6	97,73	29,45	97,33	30,12	96,77	29,05	89,02	46,32
0.7	100	19,54	99,33	19,95	100	19,14	91,75	40,56
0.8	100	13,12	100	13,38	100	12,58	92,68	32,93
0.9	100	9,77	100	9,91	100	9,91	95,43	23,43
mean	95,40	34,56	94,73	36,37	95,06	34,79	89,09	48,16

Anmerkung: Angaben in Prozent; p stellt den Cutt Off für die Klassenwahrscheinlichkeit dar

Insgesamt schneidet Random Forest hinsichtlich der Zuverlässigkeit sowie dem Anteil klassifizierter Anfragen sichtlich besser ab. Bereits ab einer Klassenwahrscheinlichkeit von $p = 0.4$ können 58% der Anfragen klassifiziert werden, wovon wiederum 90% richtig klassifiziert sind. Um die selbe Zuverlässigkeit wie bei RF zu erzielen, muss man die erforderliche Klassenwahrscheinlichkeit bei SVM auf mindestens $p = 0.7$ festlegen. Dabei werden jedoch weniger Anfragen klassifiziert (40,56% gegenüber 58%). In Anbetracht dieser Ergebnisse wird für sämtliche nachfolgenden Modellierungen der C-Parameter auf 10 und $ntree$ auf 200 festgelegt.

Da die Klassenwahrscheinlichkeiten herangezogen werden um eine möglichst zuverlässige Klassifikation zu ermöglichen, soll nun betrachtet werden wie die Klassenwahrscheinlichkeiten bezüglich der einzelnen Themen aussehen. In Abbildung 3 sind sämtliche Klassenwahrscheinlichkeiten für RF und SVM ersichtlich.

Abbildung 3: Klassenwahrscheinlichkeiten der Testklassifikation

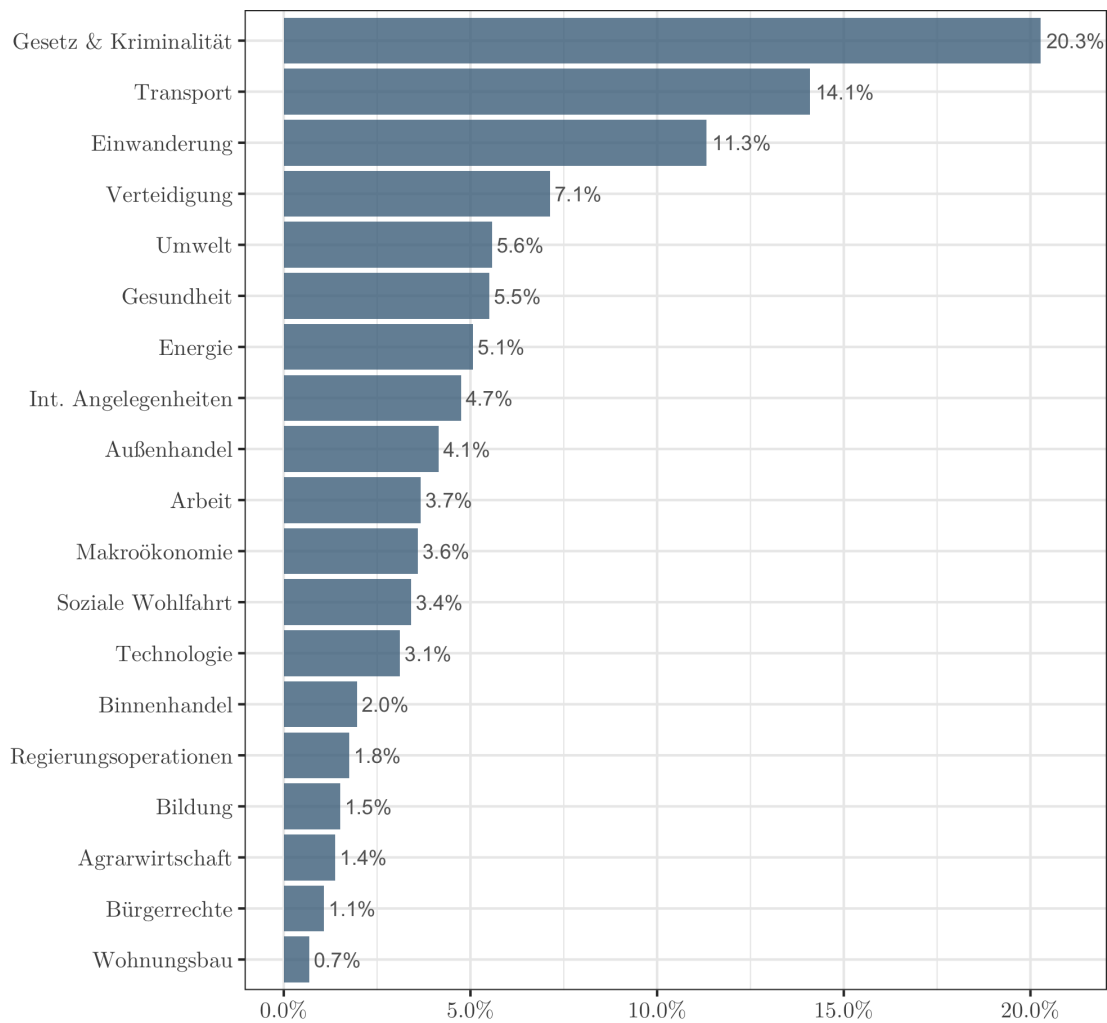


Dabei zeigt sich, dass RF bei einzelnen Themen schlechter abschneidet wie SVM. So liegen für die Themen *Wohnungsbau* und *Bürgerrechte* keinerlei Klassenwahrscheinlichkeiten vor, da durch Random Forest keinerlei Anfragen zu diesen zwei Themen zugeordnet wurden. Ausgehend der leicht besseren themenspezifischen Performance von Support Vector Machines, liegt es nahe eine Kombination von RF und SVM für die Klassifikation heranzuziehen. Hinsichtlich der Testklassifikation hat sich gezeigt, dass eine kombinierte Klassifikation von RF ($p \geq 0.5$) und SVM ($p \geq 0.75$) eine Klassifikation von über 52% ermöglicht, wovon wiederum über 93% richtig klassifiziert wurden.

In Abbildung 4 ist die Verteilung der Themen nach der finalen Klassifikation mittels einer Kombination von RF und SVM ersichtlich.¹³ Die jeweiligen Klassenwahrscheinlichkeiten können im Anhang in Abbildung A2 eingesehen werden. Letztlich konnten insgesamt 3982 der 6349 Anfragen (63%) klassifiziert werden. Hierbei sei jedoch anzumerken, dass einige der nicht klassifizierten Anfragen aufgrund starker Themenüberschneidungen auch auf manuelle Weise nur bedingt klar klassifiziert werden könnten.

¹³Die Klassifikation erfolgte dabei auf die selbe Weise wie es bei der Testklassifikation der Fall war. So wurden zunächst sämtliche Anfragen mit RF klassifiziert, für welche die Klassenwahrscheinlichkeit $p \geq 0.5$ war. Anschließend wurden die nicht-klassifizierten Anfragen durch SVM klassifiziert, für welche die Klassenwahrscheinlichkeit $p \geq 0.75$ betrug.

Abbildung 4: Verteilung der Themen nach Klassifikation (RF + SVM)

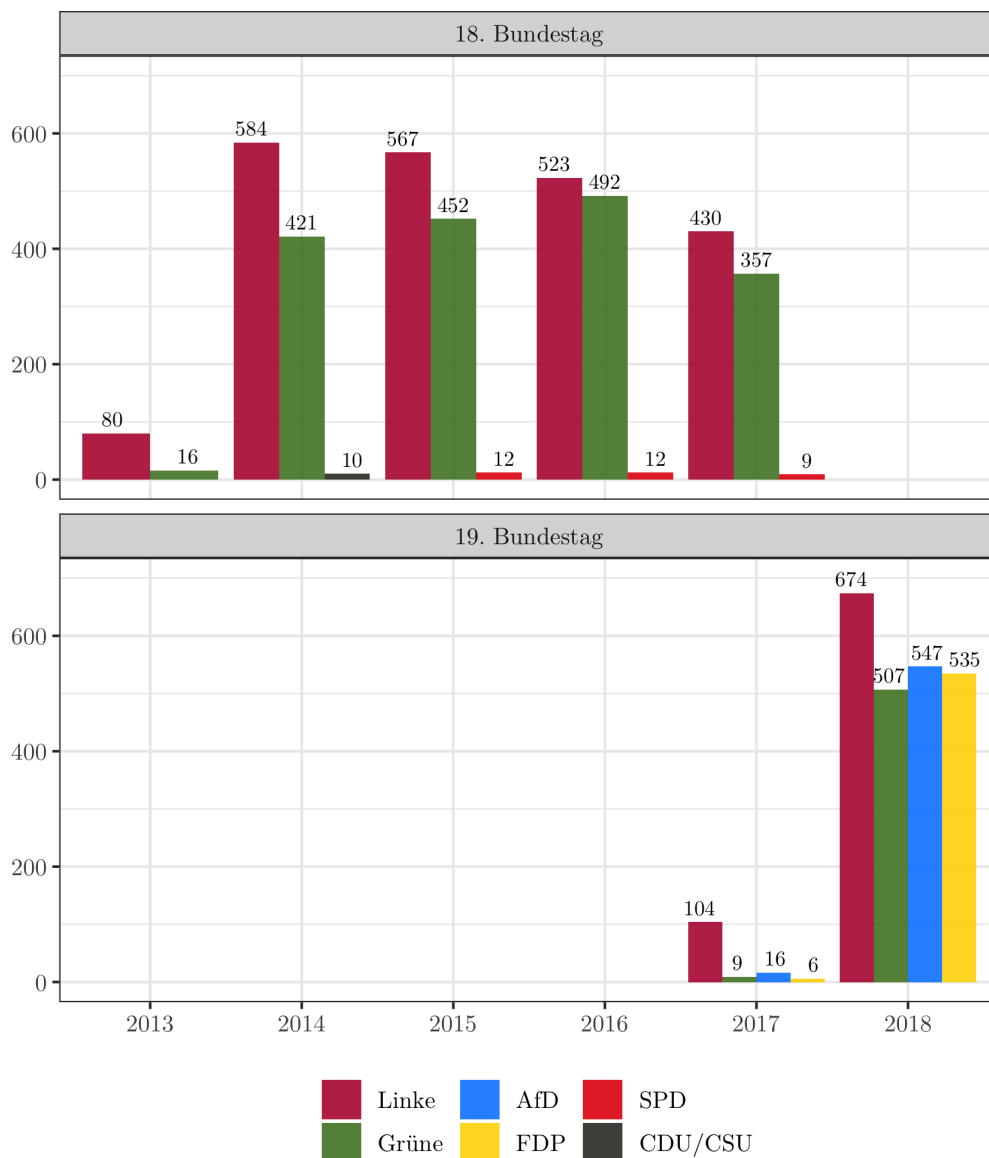


Hinsichtlich der Themenverteilung hat sich die ungleiche Verteilung leicht verstärkt, was jedoch der Tatsache geschuldet ist, dass die Genauigkeit der Klassifikation für die häufig vorkommenden Themen besser ausfällt. Summiert man die Anteile der Themen *Gesetz & Kriminalität*, *Transport*, *Einwanderung* und *Verteidigung*, kommt man bereits auf einen Anteil von über 52%. Summiert man hingegen die vier am wenigsten vorkommenden Themen *Wohnungsbau*, *Bürgerrechte*, *Agrarwirtschaft* und *Bildung*, so kommt man auf einen Anteil von weniger als 5%. Die teils besonders kleinen Anteile mancher Themen sollten bezüglich der Analyse als kritisch angesehen werden.

4.2 Deskriptive Analyse

Bei der Betrachtung der Anzahl der gestellten Anfragen (siehe Abbildung 5) wird besonders deutlich, dass es sich hierbei um ein Mittel der Opposition handelt. Zwar haben auch die Regierungsparteien Anfragen gestellt, jedoch handelt es sich bei sämtlichen dieser Anfragen um eine Anfrage bzgl. der Kriminalitätsstatistik des jeweiligen Monats. Folglich wird die CDU/CSU sowie die SPD aus den nachfolgenden Auswertungen ausgeschlossen. Die Linke hat im Vergleich zu den anderen Parteien in jedem Jahr die meisten Anfragen gestellt. Dennoch befindet sich die Anzahl gestellter Anfragen der Oppositionsparteien auf einem insgesamt ähnlich hohem Niveau.

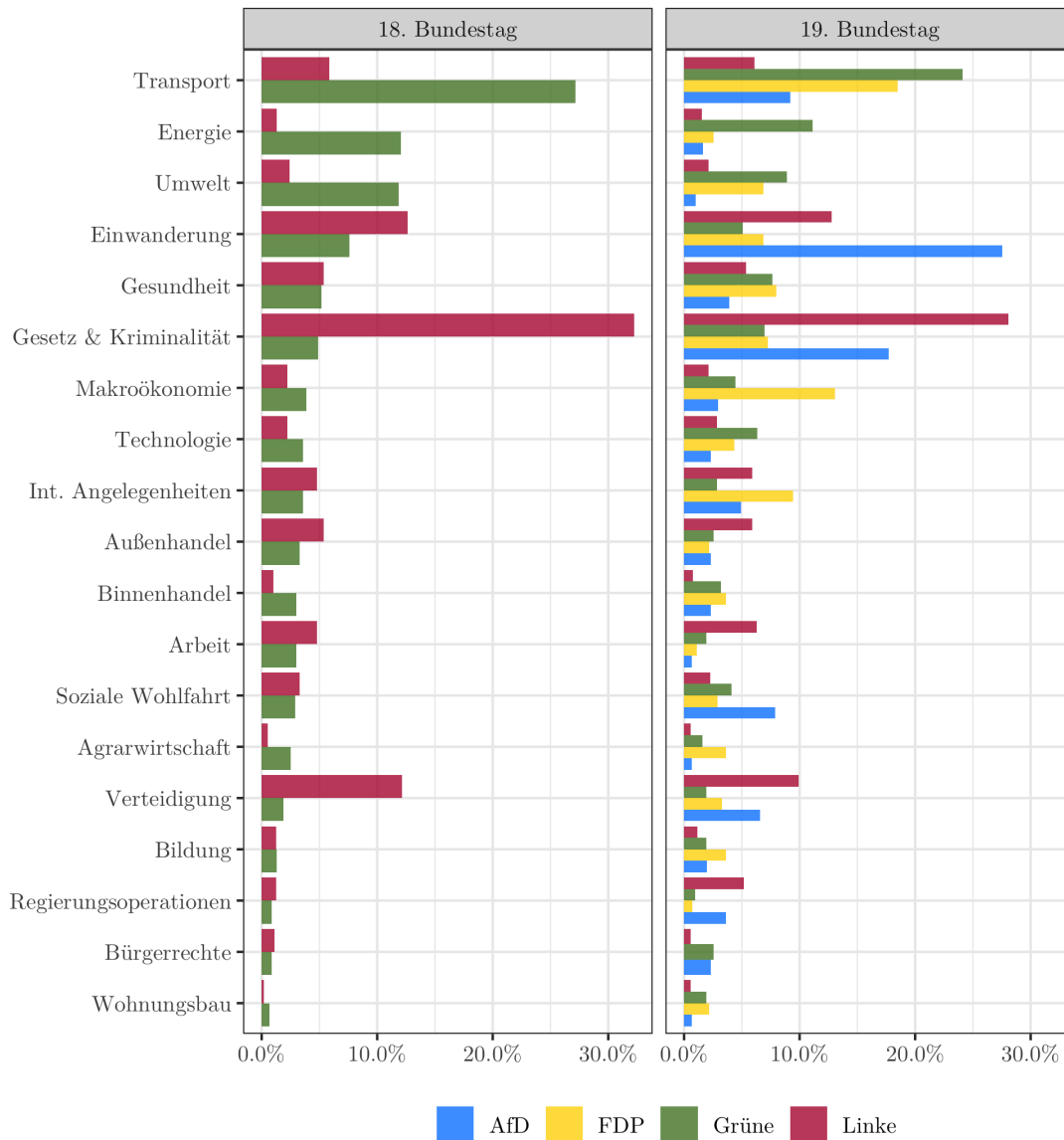
Abbildung 5: Anzahl gestellter Anfragen nach Partei



Anmerkung: Anzahl der kleinen Anfragen nach Jahr und Legislaturperiode

In Abbildung 6 sind die prozentualen Anteile der Themen nach Partei für den 18. und 19. Bundestag ersichtlich. Hierbei lassen sich parteispezifische Themenschwerpunkte identifizieren. Darüber hinaus Bei den Grünen sind vor allem die Themen *Transport*, *Umwelt* und *Energie* stark vertreten. Hinsichtlich zum Thema *Transport* sind es vor allem Bereiche¹⁴ wie E-Mobilität, größere Infrastrukturprojekte oder Schienenverkehrsthemen.

Abbildung 6: Anteil der Anfragen nach Thema und Partei



Anmerkung: Der prozentuale Anteil stellt den parteispezifischen Anteil des jeweiligen Themas dar

¹⁴Die hier genannten Unterthemen der gestellten Anfragen wurden mittels einer explorativen Sichtung der einzelnen Anfragen identifiziert. Dies gilt ebenso für die nachfolgend genannten Unterthemen. Auf den Aspekt der differenzierten Betrachtung einzelner Themen wird im Fazit noch einmal detailliert eingegangen.

Bei der Linken lassen sich ebenso klare Themenschwerpunkte identifizieren. Das Thema *Gesetz & Kriminalität* macht für den 18. Bundestag bereits einen Anteil von über 30% aller gestellten Anfragen der Linken aus. Mit jeweils ca. 12% folgen dann die Themen *Einwanderung* und *Umwelt*. Beim Thema *Gesetz & Kriminalität* liegen seitens der Linken Anfragen bzgl. rechts-extremen, islamfeindlichen und antisemitischen Straftaten vor. Im Hinblick auf das Thema *Einwanderungen* steht maßgeblich die Flüchtlingsthematik im Fokus. Mit über 27% macht das Thema *Einwanderung* bei der AfD den größten Anteil der gestellten Anfragen aus. Anschließend folgt mit knapp 18% das Thema *Gesetz & Kriminalität*. *Transport* und *Soziale Wohlfahrt* folgen mit jeweils weniger als 10%. Hinsichtlich der *Einwanderung* liegt der Schwerpunkt tendenziell auf der illegalen Einwanderung. Bei der FDP stellt das Thema *Transport* mit knapp 19% den größten Anteil dar. Anschließend folgt das Thema *Makroökonomie* mit ca. 13%. Generell stechen bei der FDP keine Themen deutlich hervor so wie es bei den anderen Parteien der Fall ist wo stets ein Thema vorhanden ist, welches mindestens 25% der gesamten Anfragen der jeweiligen Partei ausmacht.

In Anbetracht der Hypothese *H1* und den bisherigen Ausführungen zeigt sich, dass für jede Partei ein klarer Themenschwerpunkt vorliegt. Zieht man für jede Partei das Thema mit dem größten parteispezifischen Anteil heran, so ergibt sich das folgende Bild:

Bei der AfD sticht das Thema *Einwanderung* mit 27,5% deutlich hervor. Das Thema *Gesetz & Kriminalität* ist hingegen bei der Linken mit 28% deren klarer Themenschwerpunkt. *Transport* weist bei den Grünen (24%) und der FDP (18,5%) den größten Anteil ihrer Anfragen dar. Folglich kann bestätigt werden, dass die Parteien Themenschwerpunkte aufweisen. Jedoch fällt die Differenz zwischen den Parteien weniger stark aus als vermutet. Grüne und FDP haben beide das Thema *Transport* als Themenschwerpunkt. AfD und Linke weisen zwar bzgl. der größten Anteile unterschiedliche Themenschwerpunkte auf, jedoch sind beide Parteien bei beiden Themen deutlich vertreten. Ebenso sticht die FDP hervor, da diese bei keinem Thema einen Anteil von über 20% aufweist. Dafür liegt hier eine moderate Präsenz bei einigen Themen vor. Folglich kann die Hypothese *H1* als teilweise bestätigt angesehen werden.

Hypothese *H2* kann hingegen in keiner Weise bestätigt werden. Trotz des Einzugs der AfD und FDP liegt keine merkliche Veränderung hinsichtlich der parteispezifischen prozentualen Anteile der einzelnen Themen vor. Ausgehend der *Issue Ownership* sollte unter anderem der Anteil der Anfragen zum Thema *Einwanderung* bei den Parteien Linke und Grüne deutlich angestiegen sein. Bei beiden Parteien hat sich jedoch nahezu nichts verändert. Bei den Grünen ist der prozentuale Anteil sogar von 7,6% auf 5,1% gesunken. Ausgehend der Daten und den kaum vorhandenen Veränderungen bezüglich den parteispezifischen prozentualen Anteile, muss die Hypothese *H2* verworfen werden. Ein möglicher Grund für die ausbleibende Veränderung kann unter anderem in den Anfragen selbst begründet liegen. So können sich die Parteien in ihrer parlamentarischen Arbeit auf die neu hinzugekommenen Themen verstärkt fokussieren, jedoch muss sich dies nicht zwangsläufig in deren Anfragen abbilden. Auf diesen Punkt soll im Fazit nochmals detaillierter eingegangen werden.

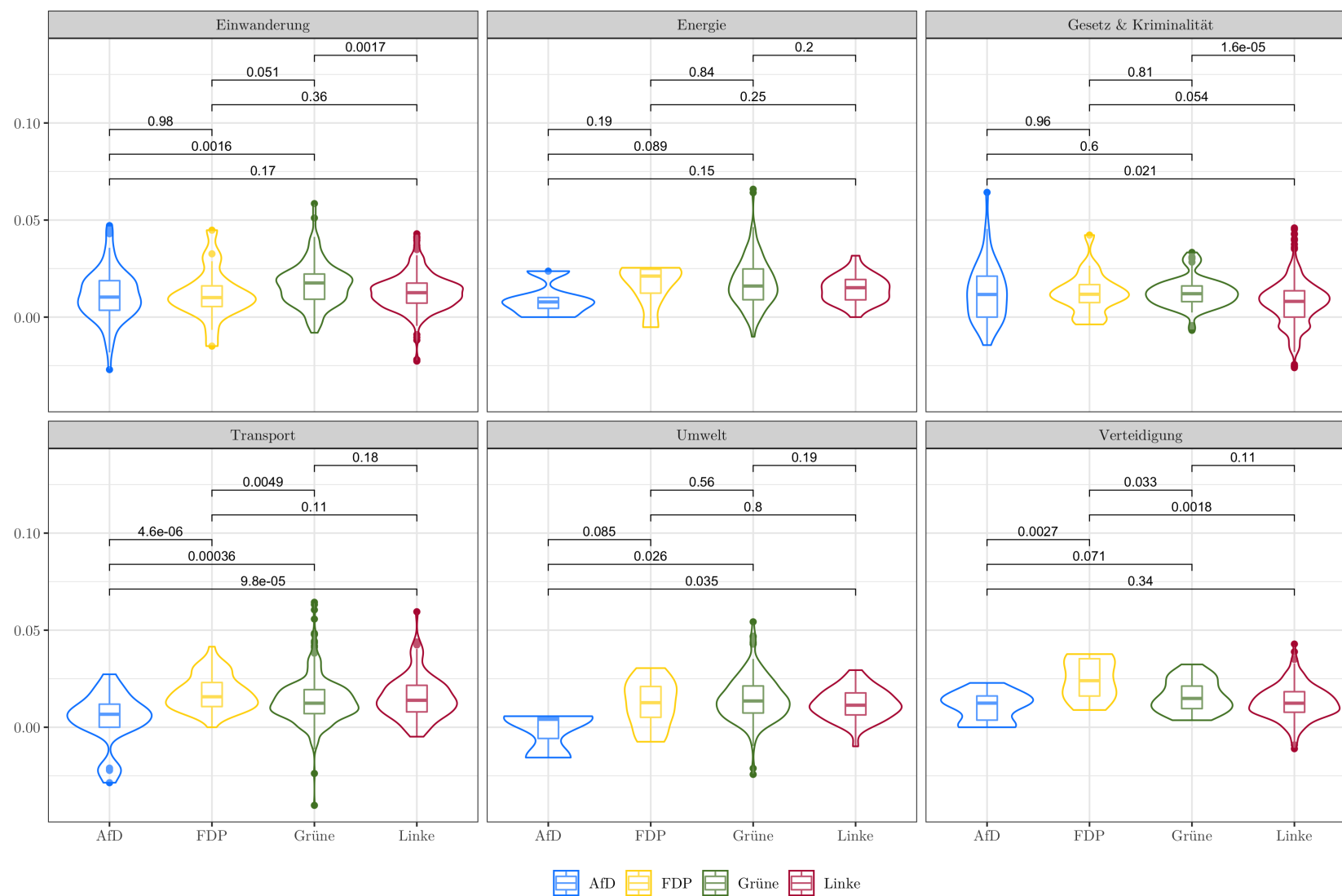
4.3 Sentiment Analyse

Auch wenn der Einzug der AfD und FDP zu keiner merklichen Veränderungen bei den prozentualen Themenanteilen von Grünen und Linken führte, kann davon ausgegangen werden, dass das Sentiment der einzelnen Parteien unterschiedlich ausfallen kann. In Anbetracht der Anzahl der Themen sollen dabei die relevantesten Themen untersucht werden, welche zugleich einen tendenziell hohen Anteil von mindestens 2 Parteien aufweisen. In Abbildung 7 sind die Sentiments der einzelnen Parteien für die Themen *Einwanderung*, *Energie*, *Gesetz & Kriminalität*, *Transport*, *Umwelt* und *Verteidigung* ersichtlich. Darüberhinaus wurde jeweils ein Mittelwertvergleich in Form eines t-Tests durchgeführt. Auf diese Weise ist es möglich signifikante Unterschiede zwischen den einzelnen Parteien feststellen zu können.

Was dabei besonders klar hervorsticht ist das tendenziell grundlegend positive Sentiment bei allen Parteien. Dies kann darin begründet liegen, dass die kleinen Anfragen primär der Informationsbeschaffung dienen, wobei zugleich auch teils das Ziel einer Aufmerksamkeitsbeschaffung verfolgt wird. Darüberhinaus findet ebenso keine Aussprache statt, wodurch die emotionale Komponenten vermutlich weniger stark ausgeprägt ist wie bei einer Rede (vgl. Kepplinger 2007: 308 f.).

Dennoch können Unterschiede zwischen den einzelnen Parteien festgestellt werden. Hinsichtlich der Mittelwertvergleiche ergeben sich entgegen der Erwartung wenig signifikante Unterschiede. Beim Thema *Einwanderung* ist das Sentiment der Grünen im Vergleich zu allen anderen Parteien signifikant positiver. Ebenso liegt kein signifikanter Unterschied zwischen der AfD und der Linken vor. Bei dem Thema *Energie* weist die AfD im Durchschnitt das negativste Sentiment auf. Zudem kann hierbei festgestellt werden, dass das Sentiment der Grünen im Durchschnitt signifikant positiver ausfällt als bei der AfD. Die Linken weisen beim Thema *Gesetz & Kriminalität* erneut ein signifikant durchschnittlich negativeres Sentiment im Vergleich zu den Grünen auf. Einige signifikante Unterschiede lassen sich beim Thema *Transport* feststellen. So hat die AfD im Durchschnitt das negativste Sentiment, welches im Vergleich zu allen Parteien einen signifikanten Unterschied darstellt. Ebenso ist das durchschnittliche Sentiment der FDP im Vergleich zu den übrigen Parteien, ausgenommen der Linken, signifikant positiver. Das Thema *Umwelt* gibt es einen erheblichen Unterschied zwischen der AfD und den übrigen Parteien. So weißt die AfD ein durchschnittlich merklich negativeres Sentiment auf als die anderen Parteien. Im Vergleich zu den übrigen Parteien fällt dieser Unterschied zudem deutlich signifikant aus. Die FDP weißt beim Thema *Verteidigung* erneut das durchschnittlich positivste Sentiment auf, welches ebenso zu allen übrigen Parteien einen signifikanten Unterschied darstellt. Die Linke hingegen weißt das negativste Sentiment auf, wobei sich dies durchschnittlich lediglich zur FDP signifikant unterscheidet.

Abbildung 7: Mittelwertvergleich (t-Test) der Sentiments für ausgewählte Themen



In Anbetracht der Ergebnisse können bei allen ausgewählten Themen signifikante Unterschiede zwischen den Parteien festgestellt werden. Zwar liegen nicht zwischen allen Parteien signifikante Unterschiede vor, jedoch ist eine Heterogenität der Sentiments durchaus ersichtlich. Überraschend ist dennoch der Aspekt, dass zwischen der Linken und der AfD bei 3 der 6 ausgewählten Themen keine signifikanten Unterschiede vorliegen. Dabei sei jedoch hervorzuheben, dass die Ergebnisse bezüglich mancher Themen aufgrund der teils sehr geringen Anzahl an Anfragen kritisch eingestuft werden sollten. So weist die AfD bei dem Thema *Umwelt* lediglich 3 Anfragen auf. Je nach Partei und Thema ergibt sich ein ähnliches Bild, wodurch die Sentiment Analyse kritisch einzustufen ist. Dennoch sticht die AfD mit ihrem durchschnittlich tendenziell negativeren Sentiment hervor. Insbesondere bei den Themen *Energie*, *Transport* und *Umwelt* zeigt sich dies.

Abseits dessen lassen sich durch eine explorativen Sichtung einiger Anfragen der verschiedenen Parteien teilweise tendenzielle unterschiedliche Schwerpunkte bezüglich der Unterthemen feststellen. So hat die Linke beim Thema *Einwanderung* ein ebenso ähnlich negatives Sentiment wie die AfD und die FDP. Dies ist jedoch unter anderem darauf zurückzuführen, dass die Linke häufig die Todesopfer von Flüchtlingen thematisiert, wohingegen die AfD die illegale Migration thematisiert. Hinsichtlich *Gesetz & Kriminalität* stellt die Linke vermehrt Anfragen zu rechtsextremen Straftaten, wohingegen die AfD vermehrt Anfragen zu linksextremen Straftaten stellt. Diese detaillierten Unterschiede wurden bei der hier durchgeführten Sentiment Analyse leider nicht berücksichtigt.

Die Hypothese *H3* kann in Hinblick der Ergebnisse als tendenziell bestätigt angesehen werden. So liegen einige signifikante Unterschiede bei den Sentiments vor und eine Heterogenität der Sentiments kann ebenso festgestellt werden, wobei diese Unterschiede nicht stets vorhanden sind. Ebenso sei darauf zu verweisen, dass die Analyse mittels inferenzstatistischer Methoden in Anbetracht der teilweise sehr geringen Fallzahlen als kritisch einzustufen ist. *H3* kann folglich nur als tendenziell bestätigt angesehen werden. Um diesen Punkt besser beurteilen zu können, sind detailliertere Analysen mit einer größeren Datengrundlage notwendig.

5 Fazit

Zusammenfassend konnte sich zeigen, dass die Parteien alle Themenschwerpunkte aufweisen. Auch wenn dabei teilweise Überschneidungen der Themenschwerpunkte vorliegt, kann eine gewisse Differenz zwischen den Oppositionsparteien festgestellt werden. Diese fallen dabei jedoch geringer aus als vermutet. Dennoch kann der erste Teil der Forschungsfrage positiv beantwortet werden. Ein hingegen unerwartetes Ergebnis ist der nicht vorhandene Einfluss der neu eingezogenen Parteien AfD und FDP auf die Themenschwerpunkte der Grünen und Linken. Trotz der teilweisen Unterschiede der Themenschwerpunkte der AfD und FDP zu den bisherigen Oppositionsparteien, haben sich die Schwerpunkte der Grünen und Linken kaum verändert. Dies kann einerseits an der nicht vorhandenen Aussprache der Anfragen liegen, wodurch ein Einfluss dieser auf andere Parteien kaum vorhanden ist. Andererseits waren die Linken und Grünen bereits vor dem Einzug der AfD und FDP bei deren Themenschwerpunkten vertreten, wodurch hier eventuell kein Handlungsbedarf gesehen wurde um die Dominanz der anderen Parteien bei gewissen Themen zu mindern. Der zweite Teil der Forschungsfrage kann folglich negativ beantwortet werden, nach dem Einzug der AfD und FDP haben sich keine Veränderungen bei Grünen und Linken aufgezeigt.

Hinsichtlich der Sentiments konnte eine Heterogenität der Sentiments festgestellt werden. So unterscheidet sich das Sentiment zwischen den Parteien oftmals signifikant, auch wenn dies zwischen den Parteien nicht konsistent der Fall ist. Dennoch gibt es hierbei ein Problem mit der Fallzahl bei manchen Themen und Parteien, wodurch die Ergebnisse als weniger stabil angesehen werden müssen. Hinzu kommt die mangelnde Differenzierung der Unterthemen. So kann zwar ein ähnliches Sentiment vorliegen, jedoch können sich die Anfragen auf völlig unterschiedliche Aspekte beziehen. Dies wird durch die hier durchgeführte Analyse leider nicht berücksichtigt, wobei eine solche Berücksichtigung schweurig umzusetzen wäre.

Generell muss in Anbetracht der Analyse in Frage gestellt werden, ob die kleinen Anfragen eine adäquate Datengrundlage für die aufgestellte Forschungsfrage und den damit einhergehenden Hypothesen darstellt. Es ist zwar durchaus nachvollziehbar, dass Parteien die kleinen Anfragen gerade bei Themen nutzen, mit denen sie sich verstärkt auseinandersetzen. Mittels einer Auswertung der Worthäufigkeiten der einzelnen Parteiprogramme (siehe Abbildung A3) kann sich lediglich bei der AfD eine tendenzielle Überschneidung der Themenschwerpunkte erkennen lassen. So kommen im Parteiprogramm der AfD häufig Wörter wie *Integration*, *Einwanderung*, *Kultur*, *Sprache* vor. Bei den anderen Parteien lassen sich durch die Worthäufigkeitsanalyse die hier identifizierten Themenschwerpunkte jedoch nicht vergleichbar wiedererkennen. Dabei sei jedoch anzumerken, dass die Worthäufigkeitsanalyse nicht ausreichend ist um mögliche Differenzen zwischen den Themenschwerpunkten in den Parteiprogrammen und den identifizierten Themenschwerpunkten anhand der kleinen Anfragen zu ermitteln. Dennoch lässt sich hierdurch eine mögliche Differenz erkennen, welche in künftigen Arbeiten analysiert werden könnte. Ebenso fehlt bei den kleinen Anfragen die Aussprache in Form einer parlamentarischen

Debatte, wodurch eine Einflussnahme auf andere Parteien nur bedingt möglich erscheint. Jedoch kann ein Wirkungsmechanismus zwischen den kleinen Anfragen und der parlamentarischen Agenda. Für eine Analyse eines solchen Wirkungsmechanismus müsste ein größerer Zeitraum mit mehr Daten betrachtet werden.

In Hinblick auf die Sentiments der einzelnen Parteien ist die Verwendung kleiner Anfragen womöglich nicht das beste Mittel. Einerseits findet bei den kleinen Anfragen keine Aussprache statt, wodurch diese tendenziell neutraler erscheinen, da es sich hierbei mehr um eine Informationsbeschaffung handelt. Andererseits ist die Vergleichbarkeit nur bedingt gegeben. Die Anfragen wurden zwar nach Themen des *Comparative Agendas Project* kodiert, jedoch gibt es bei den einzelnen Themen nochmals weitere Differenzierungsmöglichkeiten. Die Verwendung von parlamentarischen Reden der einzelnen Parteien zu den selben Themen erscheint dabei folglich als das adäquatere Mittel. Die Themen, welche auf das *Comparative Agendas Project* zurückzuführen sind, könnten folglich ebenso dazu dienen um die parlamentarischen Tagesordnungen zu kategorisieren und die Reden der einzelnen Parteimitglieder zu analysieren. Auf diese Weise ist eine bedeutend bessere Vergleichbarkeit gegeben als bei den kleinen Anfragen, wo das konkrete Thema, zu welchem die Anfrage gestellt wird, von der Fraktion selbst vorgegeben wird. Eine solche Analyse kann auch mit der hier getätigten Analyse der kleinen Anfragen kombiniert werden.

6 Literaturverzeichnis

- Ali, Jehad, Khan, Rehanullah, Ahmad, Nasir & Maqsood, Imran 2012:** Random Forests and Decision Trees, *International Journal of Computer Science Issues (IJCSI)* 9, 272.
- Ansolabehere, Stephen & Iyengar, Shanto 1994:** Riding the Wave and Claiming Ownership over Issues: The Joint Effects of Advertising and News Coverage in Campaigns, *Public Opinion Quarterly* 58, 335–57.
- Banfield, Robert E, Hall, Lawrence O, Bowyer, Kevin W & Kegelmeyer, W Philip 2007:** A Comparison of Decision Tree Ensemble Creation Techniques, *IEEE transactions on pattern analysis and machine intelligence* 29, 173–80.
- Baumgartner, Frank R, Green-Pedersen, Christoffer & Jones, Bryan D 2006:** Comparative Studies of Policy Agendas, *Journal of European Public Policy* 13, 959–74.
- Ben-Hur, Asa & Weston, Jason 2010:** A User’s Guide to Support Vector Machines, in *Data mining techniques for the life sciences*, Springer.
- Bélanger, Éric 2003:** Issue Ownership by Canadian Political Parties 1953-2001, *Canadian Journal of Political Science/Revue canadienne de science politique* 36, 539–58.
- Bélanger, Éric & Meguid, Bonnie M 2008:** Issue Salience, Issue Ownership, and Issue-Based Vote Choice, *Electoral Studies* 27, 477–91.
- Boulesteix, Anne-Laure, Janitza, Silke, Kruppa, Jochen & König, Inke R 2012:** Overview of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Bioinformatics, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2, 493–507.
- Bradford, Jeffrey P et al. 1998:** Pruning Decision Trees with Misclassification Costs, in *European conference on machine learning*, Springer.
- Breiman, Leo 1996:** Bagging Predictors, *Machine learning* 24, 123–40.
- Breiman, Leo 2001:** Random Forests, *Machine Learning* 45, 5–32.
- Breiman, Leo & Cutler, Adele 2018:** Breiman and Cutler’s Random Forests for Classification and Regression, verfügbar unter: <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>.
- Breiman, Leo, Friedman, Jerome, Olshen, Richard & Stone, Charles 1984:** *Classification and Regression Trees*. New York: Chapman & Hall/CRC.
- Chan, Jonathan Cheung-Wai & Paelinckx, Desiré 2008:** Evaluation of Random Forest and Adaboost Tree-Based Ensemble Classification and Spectral Band Selection for Ecotope

- Mapping Using Airborne Hyperspectral Imagery, *Remote Sensing of Environment* 112, 2999–3011.
- Chang, Chih-Chung & Lin, Chih-Jen 2011:** LIBSVM: A Library for Support Vector Machines, *ACM transactions on intelligent systems and technology (TIST)* 2, 27.
- Christensen, Love, Dahlberg, Stefan & Martinsson, Johan 2015:** Changes and Fluctuations in Issue Ownership: The Case of S Weden, 1979–2010, *Scandinavian Political Studies* 38, 137–57.
- Collingwood, Loren et al. 2013:** RTextTools: A Supervised Learning Package for Text Classification, *The R Journal* 5, 6–12.
- Comparative Agendas Project 2019:** CAP Topics, verfügbar unter: <https://www.comparativeagendas.net/pages/master-codebook>.
- Cortes, Corinna & Vapnik, Vladimir 1995:** Support-Vector Networks, *Machine learning* 20, 273–97.
- Cristianini, Nello & Shawe-Taylor, John 2000:** *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge university press.
- Dejaeghere, Yves & Van Erkel, Patrick 2017:** The Importance of Issue-Ownership and Salience for Voters’ Knowledge of Parties’ Issue Positions, *Electoral Studies* 46, 15–25.
- Die Welt 2017:** Alle Ergebnisse und Grafiken der Bundestagswahl im Überblick, verfügbar unter: <https://www.welt.de/politik/deutschland/article168883713/Alle-Ergebnisse-und-Grafiken-der-Bundestagswahl-im-Ueberblick.html>.
- DIP 2019:** Dokumentations- Und Informationssystem Für Parlamentarische Vorgänge, verfügbar unter: <http://dipbt.bundestag.de/dip21.web/bt>.
- Feuerriegel, Stefan & Proellocks, Nicolas 2019:** Dictionary-Based Sentiment Analysis, verfügbar unter: <https://cran.r-project.org/web/packages/SentimentAnalysis/SentimentAnalysis.pdf>.
- Friedman, Jerome, Hastie, Trevor & Tibshirani, Robert 2017:** *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics New York.
- Geys, Benny 2012:** Success and Failure in Electoral Competition: Selective Issue Emphasis Under Incomplete Issue Ownership, *Electoral Studies* 31, 406–12.
- Goldhahn, Dirk, Eckart, Thomas & Quasthoff, Uwe 2012:** Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages, in *Proceedings of the eighth international conference on language resources and evaluation (lrec’12)*
- Green, Jane & Hobolt, Sara B 2008:** Owning the Issue Agenda: Party Strategies and Vote

- Choices in British Elections, *Electoral Studies* 27, 460–76.
- Green-Pedersen, Christoffer 2007:** The Growing Importance of Issue Competition: The Changing Nature of Party Competition in Western Europe, *Political studies* 55, 607–28.
- Günther, Elisabeth, Domahidi, Emese & Quandt, Thorsten 2017:** Mediale Sichtbarkeit Der Wahlbewerberinnen Und Der Themen Der Bundestagswahl 2013., *SCM Studies in Communication and Media* 6, 262–99.
- Holian, David B 2004:** He’s Stealing My Issues! Clinton’s Crime Rhetoric and the Dynamics of Issue Ownership, *Political Behavior* 26, 95–124.
- John, Peter 2006:** The Policy Agendas Project: A Review, *Journal of European Public Policy* 13, 975–86.
- Jurka, Timothy et al. 2015:** Automatic Text Classification via Supervised Learning, verfügbar unter: <https://cran.r-project.org/web/packages/RTextTools/RTextTools.pdf>.
- Kaufmann, Karen M 2004:** Disaggregating and Reexamining Issue Ownership and Voter Choice, *Polity* 36, 283–99.
- Kepplinger, Hans Mathias 2007:** Kleine Anfragen: Funktionale Analyse einer parlamentarischen Praxis, in *Res publica semper reformanda*, Springer.
- Kleinnijenhuis, Jan & Lefevere, Jonas** Issue Ownership, Issue Convergence or Spatial Proximity? Explaining the 1998-2012 Dutch Election Results from Issue Salience and Issue Positions in the Media,
- Lefevere, Jonas, Tresch, Anke & Walgrave, Stefaan 2015:** Introduction: Issue Ownership, *West European Politics* 38, 755–60.
- Liaw, Andy, Wiener, Matthew & others 2002:** Classification and Regression by random-Forest, *R news* 2, 18–22.
- Mehta, Manish, Rissanen, Jorma, Agrawal, Rakesh & others 1995:** MDL-Based Decision Tree Pruning., in *KDD*
- Meyer, David et al. 2019:** Misc Functions of the Department of Statistics, verfügbar unter: <https://cran.r-project.org/web/packages/e1071/e1071.pdf>.
- Meyer, David, Hornik, Kurt & Feinerer, Ingo 2008:** Text Mining Infrastructure in R, *Journal of statistical software* 25, 1–54.
- Mingers, John 1989:** An Empirical Comparison of Pruning Methods for Decision Tree Induction, *Machine learning* 4, 227–43.
- Neundorff, Anja & Adams, James 2018:** The Micro-Foundations of Party Competition and Issue Ownership: The Reciprocal Effects of Citizens’ Issue Salience and Party Attachments,

- British Journal of Political Science* 48, 385–406.
- Ooms, Jeroen 2019:** Pdftools: Text Extraction, Rendering and Converting of Pdf Documents, verfügbar unter: <https://cran.r-project.org/web/packages/pdftools/pdftools.pdf>.
- Pal, Mahesh 2005:** Random Forest Classifier for Remote Sensing Classification, *International Journal of Remote Sensing* 26, 217–22.
- Petrocik, John R, Benoit, William L & Hansen, Glenn J 2003:** Issue Ownership and Presidential Campaigning, 1952–2000, *Political Science Quarterly* 118, 599–626.
- Prasad, Anantha M, Iverson, Louis R & Liaw, Andy 2006:** Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction, *Ecosystems* 9, 181–99.
- Remus, Robert, Quasthoff, Uwe & Heyer, Gerhard 2010:** SentiWS-a Publicly Available German-Language Resource for Sentiment Analysis, in *Proceedings of the seventh conference on international language resources and evaluation (Irec'10)*
- SentiWS 2019:** Leipzig Corpora Collection: SentimentWortschatz, verfügbar unter: <http://wortschatz.uni-leipzig.de/de/download>.
- Stubager, Rune 2017:** What Is Issue Ownership and How Should We Measure It?, *Political Behavior*, 1–26.
- Stubager, Rune & Seeberg, Henrik Bech 2016:** What Can a Party Say? How Parties' Communication Can Influence Voters' Issue Ownership Perceptions, *Electoral Studies* 44, 162–71.
- Stubager, Rune & Slothuus, Rune 2013:** What Are the Sources of Political Parties' Issue Ownership? Testing Four Explanations at the Individual Level, *Political Behavior* 35, 567–88.
- Tagesschau 2017:** Bundestagswahl 2017 Deutschland, verfügbar unter: <https://wahl.tagesschau.de/wahlen/2017-09-24-BT-DE/index.shtml>.
- Van Camp, Kirsten 2018:** Issue Ownership as a Determinant of Political Parties' Media Coverage, *Communications* 43, 25–45.
- Van der Brug, Wouter 2004:** Issue Ownership and Party Choice, *Electoral studies* 23, 209–33.
- Walgrave, Stefaan & De Swert, Knut 2007:** Where Does Issue Ownership Come from? From the Party or from the Media? Issue-Party Identifications in Belgium, 1991–2005, *Harvard International Journal of Press/Politics* 12, 37–67.
- Walgrave, Stefaan, Lefevere, Jonas & Tresch, Anke 2012:** The Associative Dimension of Issue Ownership, *Public Opinion Quarterly* 76, 771–82.
- Walgrave, Stefaan, Lefevere, Jonas & Tresch, Anke 2014:** The Limits of Issue Ownership

Dynamics: The Constraining Effect of Party Preference, *Journal of Elections, Public Opinion & Parties* 24, 1–19.

Wu, Ting-Fan, Lin, Chih-Jen & Weng, Ruby C 2004: Probability Estimates for Multi-Class Classification by Pairwise Coupling, *Journal of Machine Learning Research* 5, 975–1005.

7 Appendix

Tabelle A1: CAP Topics und verwendete Wörter für die Kategorisierung

Hauptthema	Subthemen	Wörter zwecks Klassifizierung
Makroökonomie	<ul style="list-style-type: none"> • Arbeitslosenquote • allg. Haushaltsprobleme • Inflation • Preise & Zinsrate • nationales Budget & Schulden; • Steuerpolitik • Industriepolitik • Preiskontrolle & Stabilisierung 	Arbeitslosigkeit, Arbeitslosenquote, Zinsen, Haushaltsschulden, Steuer, Industrie
Bürgerrechte	<ul style="list-style-type: none"> • Ethnische Minderheiten • Gruppendiskriminierung • Geschlecht & sex. Diskrimination • Altersdiskriminierung • Diskriminierung von Behinderten • Wahlrechte, Partizipation & Probleme • Meinungsfreiheit & Religion • Recht auf Privatsphäre • Aktivitäten gegen die Regierung 	Diskriminierung, Minderheit, Meinungsfreiheit, Religionsfreiheit, Partizipation, Wahlrecht, Gender, Gleichberechtigung, Gleichstellung, Öffentlichkeitsbeteiligung
Gesundheit	<ul style="list-style-type: none"> • Umfassende Gesundheitsreformen • Versicherungsreform & Kosten • Regulation der Medikamentenindustrie • Anlagenkonstruktion & Regulation • Ärztliche Haftung & Betrug • Gesundheitspersonal & Ausbildung • Gesundheitsförderung & Prävention • Kinder und Säuglinge • Geistige Krankheit & Behinderung • Langzeitpflege & Sterbenskrankheit 	Pflege, Pflegeversicherung, Krankenversicherung, Krankenkasse, Arzt, Krankheit, Impfquote, Impfung, Alkoholkonsum, Arzneimittel, Gesundheitsschutz, Medizin, Patienten

Agrarwirtschaft	<ul style="list-style-type: none">• Agrarhandel• Subventionen für die Landwirtschaft• Lebensmittelüberwachung & Sicherheit• Landwirtschaftliches Marketing• Tier- und Pflanzenkrankheiten• Schädlingsbekämpfung & Tierschutz• Fischerei	Agrar, Landwirtschaft, Landwirte, Pflanzenschutzmittel, Glyphosat
Arbeit	<ul style="list-style-type: none">• Arbeitssicherheit & Arbeitsschutz• Personalentwicklung• Angestelltenbonus• Gewerkschaften• Faire Arbeitsnormen• Jugendbeschäftigung & Kinderarbeit• Elternzeit & Kinderbetreuung• Wander- und Saisonarbeiter	Gewerkschaft, Elternzeit, Kinderbetreuung, Saisonarbeit, Arbeitsschutz, Arbeitssicherheit, Mindestlohn, Mindestlöhne, Streik, Überstunden, Arbeitszeitgesetz, Fachkräftemangel, Arbeitszeit, geringfügige Beschäftigung, befristete Beschäftigung, Arbeitsbedingungen, befristeter Beschäftigung, Leiharbeit, Arbeit auf Abruf, Arbeitsunfälle, Stellenabbau, Niedriglöhne
Bildung	<ul style="list-style-type: none">• Höhere Bildung• Grund- und weiterführende Bildung• unterprivilegierter Studenten• Berufliche Bildung• Besondere Bildung• Kunst und Geisteswissenschaften	Studium, Lehrer, Schule, Universität, Weiterbildung, Geisteswissenschaften, Hochschul

Umwelt	<ul style="list-style-type: none"> • Trinkwassersicherheit • Müllentsorgung • gefährliche Abfälle & Chemikalien • Luftverschmutzung • globale Erwärmung & Lärmbelästigung • Recycling • Umweltgefahren im Innenbereich • Arten- und Waldschutz • Umweltverschmutzung & Umweltschutz • Land- und Wassereinspeisung 	Trinkwasser, Recycling, Müll, Luftverschmutzung, Plastik, Klimawandel, Emission, Artenschutz, Stickoxide, Wasserqualität, Umwelt, Klimaschutz, Insekten, Co2- Speicherung, Grundwasser, Tierschutz, Atomtransport, Ökologisch
Energie	<ul style="list-style-type: none"> • Kernenergie & Atomkraft • Strom & Wasserkraft • Natürliche Gase & Öl • Kohle • Alternativen & erneuerbare Energien • Energieeinsparung 	Kohle, Kohleausstieg, Energie, Atomkraft, Solar, Windenergie, Erneuerbare, Atomstandort
Einwanderung	<ul style="list-style-type: none"> • Einwanderungsprobleme • Flüchtlingsprobleme 	Flüchtlinge, Asyl, Migration, Einwanderer, Einwanderung, Abschiebung, Integration, Familiennachzug, Schutzsuchende, Fluchtursachen, Duldung, Zuwanderung
Transport	<ul style="list-style-type: none"> • Massentransport & Sicherheit • Autobahnbau, Wartung & Sicherheit • Flughäfen, Flugsicherung & Sicherheit • Eisenbahnverkehr & Sicherheit • LKW- und Autotransport & Sicherheit • Maritime Fragen & Sicherheit • Infrastrukturentwicklung 	Autobahn, Schienenverkehr, Bahn, Flughafen, LKW, Maut, Mobilität, Schienen, Zugverspätung, Verkehrs- wende, Verkehrsprojekt, Radverkehr, Verkehrspolitik, Straßenbrücke, Verkehrsminister

Kriminalität	<ul style="list-style-type: none"> • Wirtschaftskriminalität • organisierte Kriminalität • Illegaler Drogenhandel & -produktion • Gerichtsverwaltung • Gefängnisse • Jugendkriminalität & Jugendjustiz • Familienprobleme • Kindesmissbrauch & Kinderpornografie • Polizei, Feuer & Waffenkontrolle • Straf- & Zivilgesetzbuch • Verbrechensverhütung & -kontrolle 	<p>Rauschmittel, Steuerhinterziehung, BND, Bundesnachrichtendienst, Gefängnis, Kindesmissbrauch, Kinderpornografie, Interpol, Kindesentführung, Straftaten, Kriminalität, Sicherungsverwahrung, Tötungsdelikte, Haftbefehl, Bundeskriminalamt, Polizei, Gewalttaten, linksextrem, rechtsextrem, Ermittlungsverfahren, Gefährder</p>
Soziale Wohlfahrt	<ul style="list-style-type: none"> • Nahrungsmittelhilfe • Armut und Armutshilfe • Programme für ältere Menschen • Unterstützung für Behinderte • Soziale Dienste • ehrenamtliche Vereinigungen 	<p>Arbeitslosengeld, Renten, Altersarmut, Pension, Armut, Tafel, Kindergeld, Jobcenter, SGB-II, Vermögensungleichheit, Einkommensungleichheit, Riesterrente, Sozialgesetzbuch, Altersvorsorge</p>
Wohnungsbau	<ul style="list-style-type: none"> • Wohnungswesen • Stadtwirtschaftliche Entwicklung • Ländliche Wohnungsprogramme • Ländliche wirtschaftliche Entwicklung • Wohnungsbauprogramme • Alten- & Behindertenwohnungen • Wohnhilfe für Obdachlose • Sekundäre Hypothekenmarkt 	<p>Mietpreis, Miete, Wohnung, Obdachlos, Wohnhilfe, Wohnheim, Brachfläche, wohnen</p>

Binnenhandel	<ul style="list-style-type: none"> • Bankensystem • Finanzinstitutverordnung • Wertpapier- & Rohstoffverordnung • Verbraucherfinanzierung • Versicherungsverordnung • Konkurs/Insolvenzen • Unternehmenszusammenschlüsse • Probleme kleiner Unternehmen • Urheberrechte & Patente • Katastrophenhilfe im Inland • Tourismus • Verbraucherschutz & – betrug • Sport- & Glücksspielverordnung 	Bank, Kartell, Finanzsektor, Verbraucherschutz, Tourismus, Patent, Urheberrecht, Verbraucherschutz, Verbraucherrecht, Fluthilfe, Wettbewerbsbeschränkung, Katastrophenschutz, Finanzmarkt, Kapitalmarkt
Verteidigung	<ul style="list-style-type: none"> • Verteidigungsallianzen • Militärische Intelligenz • Rüstungskontrolle • Waffenexport • Militär & Militärgerichte • Zivilschutz & Heimatschutz • Krieg & Auslandsoperationen 	Bundeswehr, Kriegseinsatz, Kriegseinsätze, Nato
Technologie & Kommunikation	<ul style="list-style-type: none"> • Satelliten • Telekommunikationsverordnung • Regulationen Rundfunk • Wettervorhersagen • Computerindustrie & sicherheit • allgemeine Probleme bzgl. Internet 	Internet, Breitband, Telekommunikation, GEZ, Rundfunk, Digitalisierung, Raumfahrt, Cybersicherheit
Außenhandel	<ul style="list-style-type: none"> • Streitigkeiten & Vereinbarungen • Exportförderung & Regulierung • Int. private Geschäftsinvestitionen • ausländische Investmentgesellschaft • Tarif – & Einfuhrbeschränkungen • allg. Einfuhrbestimmungen • Wechselkurs & verwandte Fragen 	Handelsabkommen, Export, Import, Zölle, Einfuhrbeschränkung, Einfuhr, Steueroasen, Freihandel, Außenhandel, Export, Waffenexport, CETA, TTIP, transatlantic trade

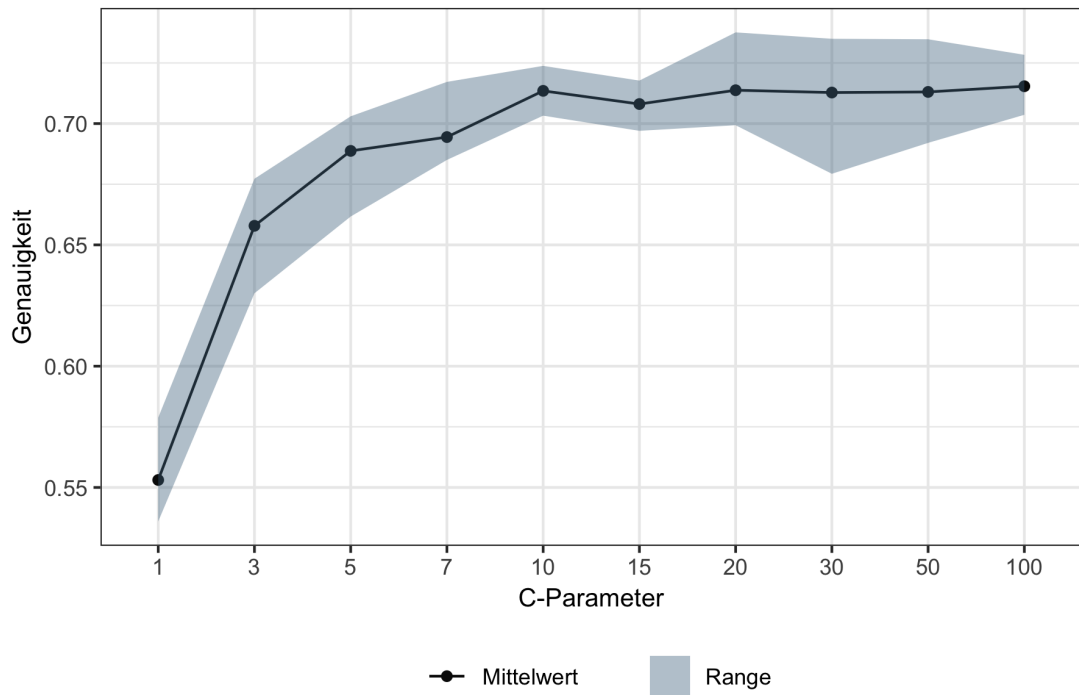
Internationale Angelegenheiten	<ul style="list-style-type: none"> • Auslandshilfe • Int. Vereinbarung (über Ressourcen) • Probleme mit Entwicklungsländern • Int. Finanzmärkte • Int. wirtschaftliche Entwicklung • Menschenrechte • Terrorismus • Probleme spezifischer Regionen 	Auslandshilfe, Entwicklungs- länder, Menschenrecht, huma- nitäre Hilfe, Brexit, islami- scher Staat, Terror, Bürger- krieg
Regierungs- operationen	<ul style="list-style-type: none"> • Zwischenstaatliche Beziehungen • Verwaltung staatlicher Immobilien • Regierungseffizienz & Kontrolle • Skandale der Regierung • Regulierung politischer Kampagnen 	Sanktionen, Bürokratieentla- stung, Bürokratieabbau, bun- desanstalt für immobilienauf- gaben, Diplomat, Diplomatie
Öffentliche Flächen	<ul style="list-style-type: none"> • Nationalparks & Erholung • Denkmäler & historische Stätten • Waldbewirtschaftung & öff. Land 	ausgeschlossen aufgrund extrem weniger Fälle
Kultur	<ul style="list-style-type: none"> • allg. kulturelle Probleme 	ausgeschlossen aufgrund extrem weniger Fälle

Anmerkung: Es wurden die Themen vom *Copmarative Agendas Project* übersetzt (Quelle: Comparative Agendas Project 2019).

Tabelle A2: Themenspezifischer Fehler nach n_{tree} (Random Forest)

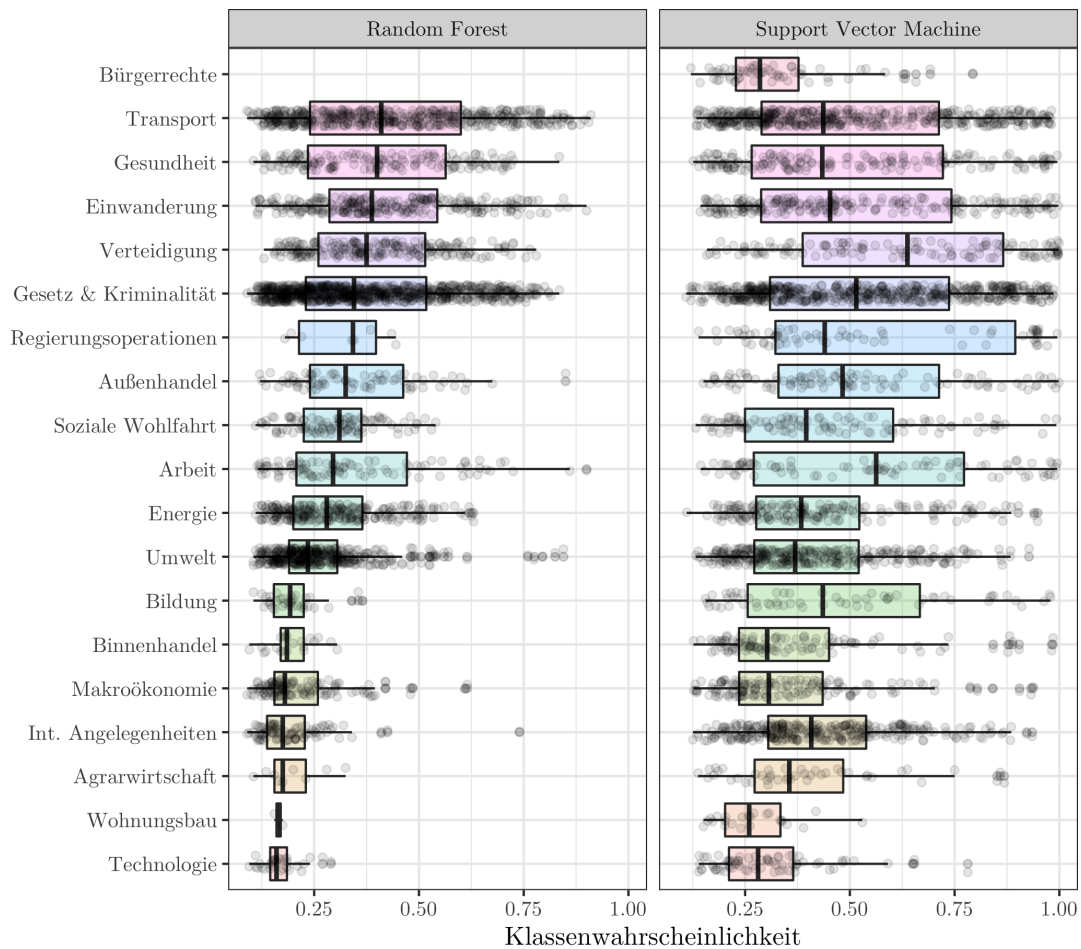
Thema	Fehler RF_{200}	Fehler RF_{500}	Fehler RF_{1500}	$N_{Anfragen}$
Gesetz & Kriminalität	5.88%	5.64%	5.88%	535
Transport	9.59%	9.59%	9.96%	348
Verteidigung	10.24%	9.64%	9.04%	206
Einwanderung	12.35%	9.96%	11.95%	347
Außenhandel	15.38%	13.46%	14.42%	137
Gesundheit	21.55%	19.83%	18.97%	159
Energie	27.91%	25.58%	25.58%	167
Arbeit	28.95%	27.63%	31.58%	114
Soziale Wohlfahrt	38.71%	29.03%	27.96%	121
Umwelt	40.4%	36.42%	33.11%	197
Makroökonomie	42.42%	41.41%	41.41%	135
Bildung	53.66%	58.54%	56.1%	50
Technologie	61.22%	61.22%	59.18%	123
Int. Angelegenheiten	66.39%	67.21%	68.03%	170
Regierungsoperationen	82.35%	73.53%	73.53%	47
Binnenhandel	84.31%	78.43%	84.31%	71
Wohnungsbau	91.67%	91.67%	95.83%	27
Agrarwirtschaft	91.89%	94.59%	94.59%	51
Bürgerrechte	100%	100%	100%	42

Abbildung A1: Genauigkeit in Abhängigkeit von C (SVM)



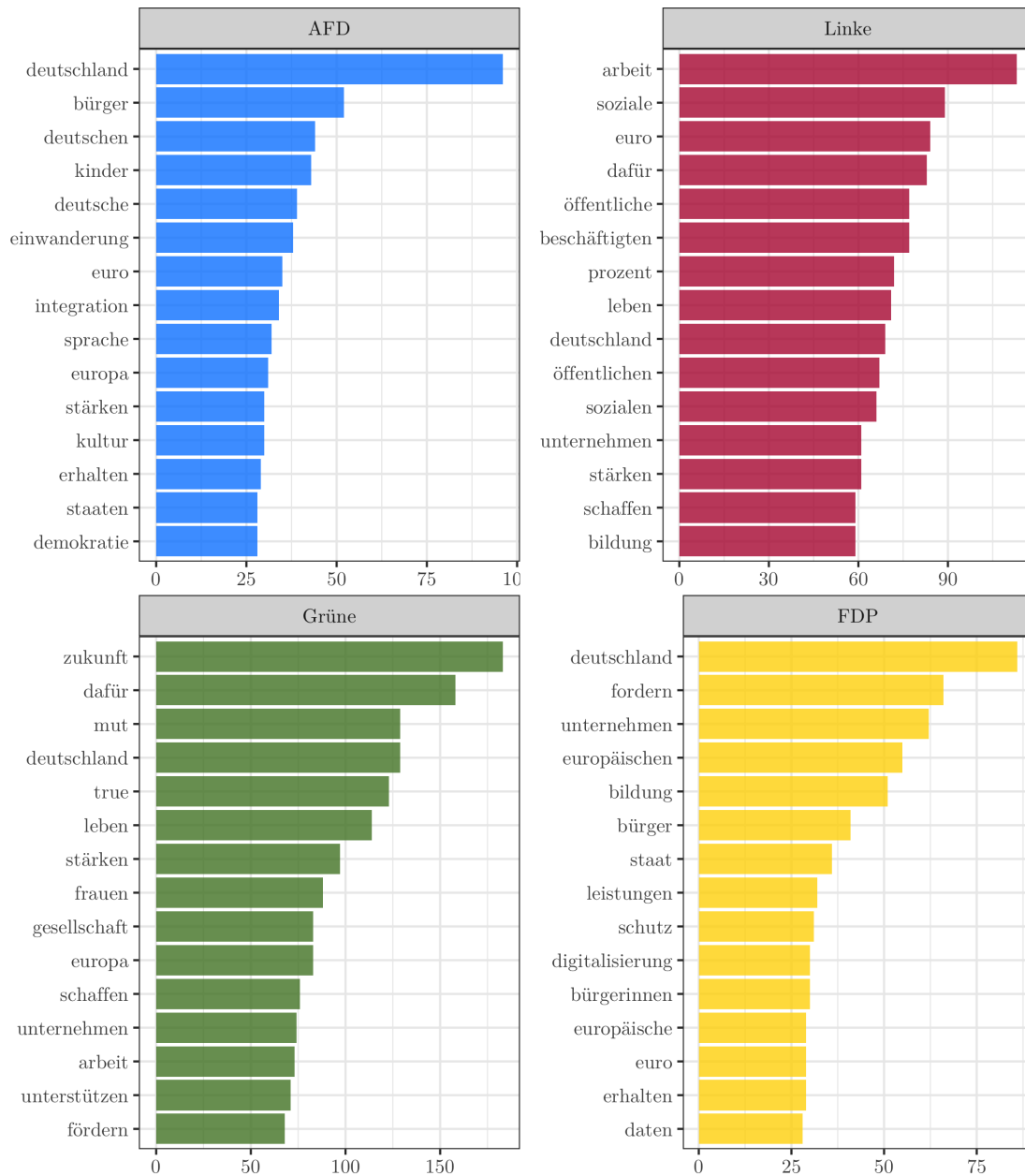
Anmerkung: Für jeden C-Wert wurden jeweils 5 Iterationen durchgeführt

Abbildung A2: Klassenwahrscheinlichkeiten (vollständige Klassifikation)



Anmerkung: ntree = 200; C-Parameter = 10

Abbildung A3: Worthäufigkeiten der Parteiprogramme



Anmerkung: Es handelt sich um die Worthäufigkeiten der Bundestagswahlprogramme aus dem Jahr 2017. Dargestellt sind die 15 am häufigsten vorkommenen Wörter. Stoppwörter und die Parteinaamen wurden dabei ausgeschlossen.

Eigenständigkeitserklärung

Hiermit versichern wir, dass wir die vorliegende Hausarbeit selbständig und nur mit den angegebenen Hilfsmitteln verfasst haben. Alle Passagen, die wir wörtlich als auch sinngemäß aus der Literatur oder aus anderen Quellen wie z. B. Internetseiten entnommen haben, sind deutlich als Zitat mit Angabe der Quelle kenntlich gemacht.

Stuttgart, 20.03.2019