

1 IDEOLOGIE EN CLASSIFICATIE IN DE HANDELINGEN
2 VAN DE TWEEDE KAMER
3
4 SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF
5 BACHELOR OF SCIENCE
6
7 JASPER VAN DER HEIDE
8 10732721
9
10 BACHELOR INFORMATIEKUNDE
11 FACULTEIT DER NATUURWETENSCHAPPEN, WISKUNDE EN
INFORMATICA
UNIVERSITEIT VAN AMSTERDAM
2018-06-28

12

	Begeleider	Tweede lezer
Titel, Naam	Dr Maarten Marx	
Affiliatie	UvA, FNWI, IvI	
Email	maartenmarx@uva.nl	.



14	Inhoudsopgave	
15	1 Introductie	3
16	2 Gerelateerd werk	4
17	2.1 Tekstclassificatie van parlementaire teksten	4
18	2.2 Classificatiemethoden	5
19	2.3 Invloed van partijnamen of sprekersnamen	6
20	2.4 Invloed van oppositie of regering	6
21	3 Methodologie	6
22	3.1 De data	6
23	3.2 Methoden	9
24	3.2.1 Deelvraag 1	9
25	3.2.2 Deelvraag 2	11
26	3.2.3 Deelvraag 3	11
27	3.2.4 Deelvraag 4	13
28	4 Evaluatie	14
29	4.1 Resultaten	14
30	4.1.1 Deelvraag 1	14
31	4.1.2 Deelvraag 2	16
32	4.1.3 Deelvraag 3	17
33	4.2 Discussie	18
34	4.2.1 Deelvraag 1	18
35	4.2.2 Deelvraag 2	18
36	4.2.3 Deelvraag 4	18
37	5 Conclusies	19
38	A Slides	19

39

Samenvatting

40

1 Introductie

Teksten van politieke partijen kunnen dienen als bron voor het bepalen van ideologische positie van andere teksten, aangezien zij zowel tekst hebben als ook een bekende ideologie in de vorm van een partij. Deze informatie kan vervolgens toegepast worden bij andere teksten die wellicht ideologisch van aard zijn. Bijvoorbeeld, aan de hand van deze informatie kan men teksten uit kranten classificeren op basis van ideologie [1, 2].

In diverse landen zijn al verschillende onderzoeken gedaan naar het classificeren van partij-affiliatie op basis van teksten van politici [3, 1]. Met deze tekstclassificatie naar partij-affiliatie proberen onderzoekers uit te vinden in hoeverre ideologie terug te vinden is in teksten van politici. De resultaten van de tekstclassificaties zijn in alle gevallen ruim boven de baseline. Maar diverse onderzoeken wijzen ook naar redenen dat dit niet alleen het gevolg is van ideologie. De resultaten van Hirst et al. [2] suggereren dat de partij-status (oppositie tegenover regering) van invloed is op de classificatie. Daarnaast laat dit onderzoek ook zien dat de partijnamen belangrijk zijn in de classificatie.

Een onderzoek gericht op het Nederlandse parlement is niet gevonden. Ook beperken veel onderzoeken zich vaak tot één classificatiemethode.

Dit onderzoek richt zich daarom op een breder scala aan mogelijke methoden en daarnaast specifiek op de Nederlandse politiek. De onderzoeksvraag luidt daarom dus ook: "In hoeverre is classificatie op basis van partij-affiliatie aan de hand van spreekbeurten in de Tweede Kamer het gevolg van ideologie?"

Deze vraag wordt beantwoord door de antwoorden te vinden op de volgende deelvragen:

1. Wat is het beste classificatiemodel voor classificatie van partij-affiliatie in de Tweede Kamer en wat is het resultaat van dit model?
2. In hoeverre is deze classificatie afhankelijk van partijnamen en namen van Kamerleden?
3. In hoeverre wordt deze classificatie bepaald door partij-status (d.w.z. oppositie of regering)?
4. In hoeverre wordt deze classificatie bepaald door links/rechts verdeling?

Daarom zal eerst gekeken worden naar classificatiemethoden en resultaten in vergelijkbare onderzoeken. Van deze classificatiemethoden zullen een aantal toegepast worden op teksten van de Tweede Kamer. Vervolgens zal door middel van de overige deelvragen bepaald worden in hoeverre dit een reflectie is van ideologie.

Overzicht van scriptie Sectie 2 bevat gerelateerd werk, met name vergelijkbare onderzoeken in andere landen. Sectie 3 bevat de methodologie van de verschillende deelvragen. Sectie 4 bevat de resultaten. Sectie 5 bevat de evaluatie van zowel de resultaten als de gehanteerde methodologie. Sectie 6 bevat ten slotte het antwoord op de onderzoeksvraag.

2 Gerelateerd werk

Toespraken in parlementen worden veel gebruikt in tekstclassificatie, omdat deze veel nette tekst bevatten en vaak gelabeld zijn. Labels zijn bijvoorbeeld naam en partij van de spreker, maar ook daar uit afleidbare labels zoals geslacht, leeftijd en partij-status (oppositie of regering).

In dit hoofdstuk zullen verschillende onderzoeken behandeld worden die tekstclassificatie hebben toegepast op parlementaire teksten. Eerst zullen de onderzoeken algemeen besproken worden. Vervolgens zal uitgebreider gekeken worden naar de effecten van verschillende classificatiemethoden. In de latere secties zullen specifieke aspecten van onderzoeken verder besproken worden.

2.1 Tekstclassificatie van parlementaire teksten

Diermeier et al. deden onderzoek naar het classificeren op basis van ideologische positie in de Amerikaanse Senaat[4]. Ze trainden hun classificatie op de speeches van de 25 meest liberale en de 25 meest conservatieve senatoren van het 101e tot en met het 107e Congres en testten op dezelfde categorieën van het 108e Congres. Een document was in dit onderzoek de verzameling van alle speeches van een senator in een congres. Deze classificatie resulteerde uiteindelijk in een nauwkeurigheid van 94% (baseline van 50%). Van de 50 senatoren in de test set, kwamen er 44 al voor in de

Later in het onderzoek vergeleken ze ook de 25 gematigd conservatieve en de 25 gematigd liberale senatoren van dezelfde congressen. Het resultaat hiervan was 52% (baseline van 50%), dus nauwelijks beter dan gokken. Als verklaring voor dit verschil ten opzichte van de uitersten zeggen ze dat gematigden een minder duidelijke ideologie hebben.

Het onderzoek van Yu et al. [5] richtte zich vervolgens op zowel het Amerikaanse Huis van Afgevaardigden als de Senaat in 2005. Een document was in dit onderzoek de verzameling van alle speeches van een senator in een Congres en het label de partij. Voor het Huis van Afgevaardigden vonden ze een nauwkeurigheid van 80.1% (baseline van 51.5%) en voor de senaat 86.0 % (baseline van 55.0%). Ze testten hun classificaties ook op de andere kamer. Van Huis van Afgevaardigden naar senaat leverde dit een nauwkeurigheid op van 88.0% (baseline van 55.0%) en andersom 67.6% (baseline van 51.5%). Hun verklaring voor dit verschil is dat het Huis van Afgevaardigden meer partisan is.

Vervolgens herhaalden ze de classificaties op het huis uit 2015, maar testten ditmaal op de senaat elk jaar tussen 1989 en 2006 afzonderlijk. Hier zien zij een stijging in nauwkeurigheid van 60% (baseline van 55.0%) in 1989 naar 87.0% (baseline van 55.0%) in 2006, maar met twee duidelijke dalen. Ze presenteren twee mogelijke verklaringen voor de trend; het veranderen van de onderwerpen en het meer partisan worden van het congres.

Als een vervolg op deze onderzoeken deden Graeme Hirst et al. een vergelijkbaar onderzoek naar het Canadese Parlement [2]. Hierbij werd zowel gekeken naar de Engelse als Franse teksten. Een document werd hier gezien als de samenvoeging van alle spreekbeurten van een spreker. Afhankelijk van taal en dataset vinden zij in dit onderzoek nauwkeurigheden van 83.2% en hoger (baseline van 65.5%).

Het onderzoek bevat ook een classificatie van het Europees Parlement. Hierbij voegen ze alle teksten van een parlamentslid bij elkaar en delen die op in

stukken van gelijke grootte. Zij vinden voor documentgrootte van 267 woorden een nauwkeurigheid van 44.0% oplopend tot 61.8% (baseline van 38-39%) voor documentgrootte van 6666.

Het onderzoek van Bhand et al. richtte zich op het classificeren van leden van het Amerikaanse congres in 2005, op basis van affiliatie (Republikeins of Democratisch)[6]. Een document hierbij was in tegenstelling tot eerdergenoemde onderzoeken een speech. Zij vonden hiervoor uiteindelijk een F_1 score van 0.68 (baseline niet vermeld).

Ferreira probeerde interventies van politici te classificeren op basis van geslacht, leeftijdsgroep, partij-affiliatie en oriëntatie in het Portugese parlement [3]. In alle classificaties kon men een F_1 score van 0.87 of hoger bereiken.

In het onderzoek van Høyland et al. werd een classificatiemodel voor partij-affiliatie op basis van teksten getraind op het vijfde Europese Parlement (1999-2004) en getest op het zesde Europese Parlement[7]. Hier verkregen zij een *macro average* F_1 score van 0.464.

2.2 Classificatiemethoden

In het onderzoek van Diermeier et al. [4] werd gebruik gemaakt van Support Vector Machines. Verder maakten ze gebruik van *tf-idf* met een minimale woordfrequentie van 50 en een minimale documentfrequentie van 10 en *Part-Of-Speech tagging*.

Het onderzoek van Yu et al. [5] maakte gebruik van Support Vector Machines en Naive Bayes, waarvan de varianten multinomial en Bernoulli. De features waren unigrams, met minimale woordfrequentie van drie en de top 50 meest voorkomende woorden weggelaten. Voor de wegingen van de features bij Support Vector Machines werd geëxperimenteerd met *boolean*, *tf-norm* en *tf-idf*. Het beste resultaat was afhankelijk van welke kamer Voor het huis van afgevaardigden was het Support Vector Machines met als weging *tf-idf* en voor de senaat Bernoulli Naive Bayes.

In het onderzoek van Graeme Hirst et al. maakten ze gebruik van support vector machines[2]. Ze experimenteerden met verschillende vormen van pre-processing, inclusief stemmen en het verwijderen van woorden op basis van te hoge of te lage frequentie. Deze variaties maakten in hun onderzoek geen grote verschillen en uiteindelijk is gekozen voor het niet stemmen, het weglaten van woorden die in minder dan vijf documenten voorkomen en resultaten van zowel met als zonder de top 500 meest frequente woorden. Daarnaast werd geëxperimenteerd met vier wegingen voor woorden: *boolean*, *tf*, *tf-norm* en *tf-idf*, waarvan *tf-idf* het beste resultaat opleverde.

In het onderzoek van Bhand et al. gebruikten ze verschillende n-grams, inclusief verschillende manieren van *smoothing*[6]. Ze testten als weging voor features zowel *boolean* als *tf*, waarbij ze vonden concludeerden dat *boolean* betere resultaten opleverden. Voor classificatiemodel experimenteerden ze met SVM en Naive Bayes. Voor het selecteren van *features* experimenteerden ze met een minimale frequentie en selectie van woorden op basis van hoogste mutual information. Uiteindelijk was het beste model bij hen een SVM met uni- en bigrams en geselecteerd op basis van mutual information.

In het onderzoek van Ferreira werd gebruik gemaakt van twee classificatiemethoden: Logistische regressie en MIRA[3]. Logistische regressie werd aangevuld met *group Lasso* regularisatie. Voor wegingen van woorden werd

177 geëxperimenteerd met *tf*, *tf-idf*, Δ -*tf-idf* en Δ -*BM-25*. Daarnaast wordt er ge-
178 bruik gemaakt van woordclustering, *Concise Semantic Analysis* en stylome-
179 trische eigenschappen. Op *Part-Of-Speech tagging* na hadden stylometrische
180 eigenschappen een duidelijke negatieve invloed op de classificatie.

181 In het onderzoek van Høyland et al. werd gebruik gemaakt van een multi
182 class support vector machine[7]. Als beste waarde voor de regularisatieterm,
183 de C-parameter, vonden zij 0.8. Daarnaast gebruikten zij *dependency disambi-*
184 *guated stems* wat bij hen een F_1 score van twee procent hoger opleverden dan
185 normale stemming.

186 2.3 Invloed van partijnamen of sprekersnamen

187 In het onderzoek van Diermeier et al. zijn de namen van de sprekers weggelaten
188 en verwijzingen naar staten die de senatoren representeren, omdat deze volgens
189 hen de classificatie te makkelijk zouden maken [4]. Hirst et al. vinden inderdaad
190 dat partijnamen (en het weglaten daarvan) bij het Europees Parlement een grote
191 invloed hebben op de classificatie [2]. Bij het Europees Parlement zien zij met
192 name het gebruik van de eigen partijnaam door een spreker, terwijl zij in het
193 Canadese parlement vooral zien dat de naam van de andere partij gebruikt
194 wordt door een spreker.

195 2.4 Invloed van oppositie of regering

196 Graeme Hirst et al. vonden in hun onderzoek dat de classificatie van spreker in
197 het Canadese parlement op basis van partij-affiliatie meer zegt over de status
198 van de partij (regering of oppositie).[2] Zo vergeleken zij de top tien karakteris-
199 tieke woorden van de liberalen en conservatieven in het 36e parlement (liberalen
200 in de regering) en het 39e parlement (conservatieven in de regering. Hier vonden
201 zij dat vier van de tien woorden van de liberalen (regering) in het 36e parlement
202 bij het 39e parlement bij de conservatieven (regering) te vinden waren. An-
203 dersom gebeurde hetzelfde met één van de tien woorden van de conservatieven
204 (oppositie) in het 36e parlement naar liberalen (oppositie) in het 39e parlement.

205 In hetzelfde onderzoek trainden ze ook hun classifiers op het ene parle-
206 ment en testten deze op het andere parlement. Hierbij vonden zij in beide
207 gevallen een nauwkeurigheid ver onder de baseline. Daarnaast deden ze ook
208 nog een classificatie op de sprekers die in beide parlementen zaten en een an-
209 dere classificatie op sprekers die niet in beide parlementen zaten. Bij de eerste
210 classificatie vonden ze nauwkeurigheden rond de baseline, terwijl in de tweede
211 situatie nauwkeurigheden gevonden werden ver boven de baseline.

212 Deze resultaten leidden de onderzoekers tot de conclusie dat de classificatie
213 voornamelijk het gevolg is van de status van de partij en minder van ideologie.

214 3 Methodologie

215 3.1 De data

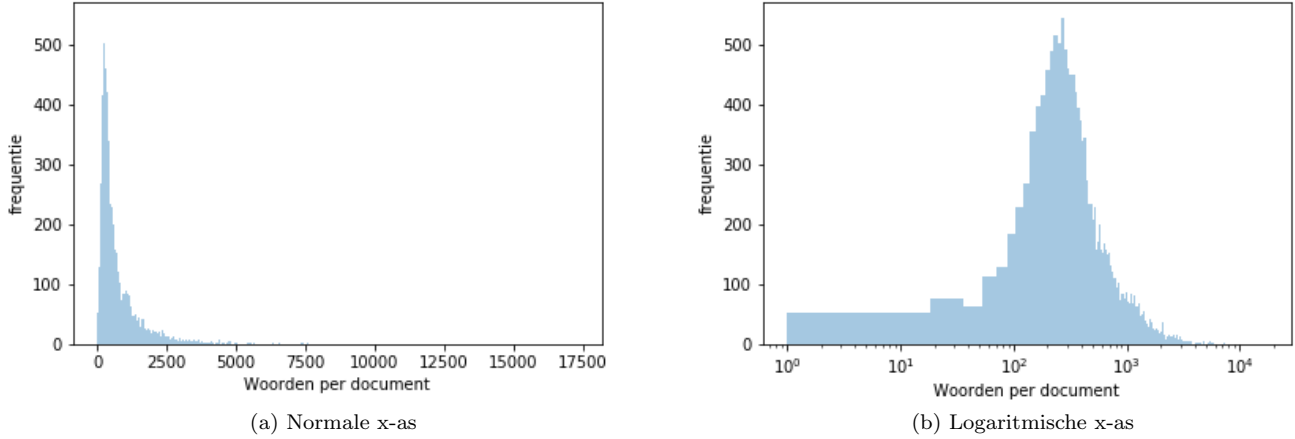
216 De data die gebruikt worden, zijn de Handelingen van de Tweede Kamer gedu-
217 rende het missionaire kabinet-Rutte II (5 november 2012 tot 22 maart 2017).
218 Er is gekozen voor dit kabinet, omdat de data hiervoor makkelijk verkrijgbaar

219 was, het kabinet lang zat - waardoor er veel data is - en het recent is waar-
220 door het makkelijker te interpreteren is. Deze data zijn in xml-formaat van de
221 website officiële bekendmakingen.nl gehaald samen met corresponderende meta-
222 data xml-bestanden. De bestanden van de Handelingen bevatten voornamelijk
223 informatie over spreekbeurten tijdens een debat, waaronder naam van een spre-
224 ker, partij-affiliatie, inhoud van de spreekbeurt en het soort spreekbeurt. Deze
225 gegevens zijn samengevoegd tot één tabel.

226 Deze dataset bestaat uit een aantal soorten spreekbeurten; debat bijdra-
227 gen, interrupties en antwoorden. Debat bijdrage is de eerste onafgebroken
228 spreekbeurt die een spreker geeft achter het spreekgestoelte, aangeduid in de
229 xml-file met het attribuut *nieuw="ja"*. Dit kan een bijdrage in een debat zijn
230 of een vraag tijdens een vragenuur. Interrupties zijn de vragen die andere po-
231 litici stellen vanachter de interruptiemicrofoon aan de spreker. De antwoorden
232 zijn vervolgens de reactie van een spreker achter het spreekgestoelte op een in-
233 terruptie. Aangezien een debat bijdrage geïnterrupteerd kan worden, kan deze
234 inhoudelijk doorlopen in een antwoord van een spreker. Er is in dit onderzoek
235 ervoor gekozen om gebruik te maken van een debat bijdrage samengevoegd tot
236 één document met alle bijbehorende antwoorden van die spreker. Daarnaast zijn
237 er verschillende soorten sprekers; de voorzitter, Tweede Kamerleden, leden van
238 het kabinet en gastsprekers. Daarnaast is alleen gekozen voor sprekers waarvan
239 er een partij-affiliatie vermeld staat, dit is niet het geval voor leden van het
240 kabinet, de voorzitter en gastsprekers (met uitzondering van Nederlandse leden
241 van het Europees Parlement).

242 Deze dataset bevat vervolgens naast de verkozen partijen van de 2012
243 Tweede Kamerverkiezingen, ook afsplitsingen van die partijen (tien in totaal) en
244 bezoeken van vertegenwoordigingen van Nederlandse partijen uit het Europees
245 Parlement (tien in totaal). Omdat van beide categorieën relatief weinig data is
246 en er overlap zit met hun oorspronkelijke partij, zijn deze er uit gehaald.

247 De documenten verschillen vervolgens in grootte. De distributie lijkt op
248 een lognormale verdeling, maar met een Kolmogorov-Smirnov test is hier geen
249 bewijs voor gevonden [8].



Figuur 1: Aantal woorden per document

Om toch de uitschieters er uit te halen, is aangenomen dat het wel lognormaal verdeeld is en zijn daarmee de documenten buiten het betrouwbaarheidsinterval van 95% eruit gehaald. De documenten met een lengte van minimaal 28 en maximaal 1492 woorden bleven daarmee over. Het gemiddelde is daarna 498 woorden en de mediaan is 386 woorden. Een totaal aantal documenten van 14899 blijven vervolgens over.

Tabel 1: Aantal documenten per partij gedurende het missionaire kabinet-Rutte II.

	Totaal	Vragenuur	Debat
SP	2284	107	2177
CDA	1901	88	1813
D66	1889	133	1756
PvdA	1821	112	1709
PVV	1700	49	1651
VVD	1694	76	1618
ChristenUnie	1068	32	1036
GroenLinks	1068	47	1021
SGP	655	10	645
PvdD	432	14	418
50PLUS	387	12	375

Deze 14899 documenten zijn verdeeld over 2984 debatten, waarbij elke vraag tijdens het vragenuur als één debat gezien wordt. Op basis van de aantallen is er voor classificatie een baseline *accuracy* van 0.15 (door altijd grootste partij te kiezen) en baseline F_1 score van 0.11 (door willekeurig te voorspellen gewogen bij aantal spreekbeurten in klasse).

261 3.2 Methoden

262 3.2.1 Deelvraag 1

263 Om deze deelvraag te beantwoorden zullen een aantal classificatiemethoden ver-
264 geleken worden. Aangezien het onmogelijk is om alle classificatiemethoden te
265 vergelijken, beperkt dit onderzoek zich tot classificatiemethoden die gebruikt
266 zijn in vergelijkbare onderzoeken, zoals besproken in 2.2. Er is ervoor gekozen
267 om alleen gebruik te maken van methoden waarvan reeds implementaties be-
268 schikbaar waren in Python. Voor alle methoden wordt gezocht naar de beste
269 parameters; een grid search. Deze grid search wordt gedaan door middel van
270 5-fold cross-validation, waarbij de trainings set steeds 80% is en de test set 20%
271 van de totale dataset.

272 **Pre-processing** Voor pre-processing is gebruik gemaakt van tokenisation en
273 lowercasing. Voor tokenisation is de reguliere expressie
274 `w+` gebruikt, waardoor alles behalve letters en cijfers weggehaald wordt. Ver-
275 volgens is er gevarieerd tussen wel of geen gebruik maken van stemming. In
276 het geval van stemming is gebruik gemaakt van de Snowball Stemmer via de
277 Python NLTK module.

278 **Bag-of-words model** Bag-of-words model is de meest gebruikte representa-
279 tie van data in vergelijkbare onderzoeken. Bij het bag-of-words model wordt
280 elk document gerepresenteerd door een vector, waarbij elke kolom een woord
281 voorstelt met een bijbehorende waarde. Voornaamste beperking van dit model
282 is dat het geen rekening houdt met de volgorde van woorden, wat een groot
283 effect kan hebben op de betekenis van een document.

284 Voor dit onderzoek zijn de volgende wegeningen voor woorden getest: *boolean*
285 (wel of niet aanwezig), *tf* (woordfrequentie), *tf-norm* (woordfrequentie genor-
286 maliseerd door documentlengte) en *tf-idf*. Daarnaast wordt in dit onderzoek
287 geëxperimenteerd met een minimale of maximale woord- of documentfrequentie.
288 Ook is gekeken naar het effect van combinaties van n-grams; unigrams, bigrams
289 en trigrams. N-grams zijn combinaties van N aantal opeenvolgende woorden.
290 Bij een unigram is elke feature gewoon één woord, terwijl bij een bigram dit
291 twee opvolgende woorden zijn. Dit kan nuttig zijn, want als bijvoorbeeld het
292 woord *asfalt* er in voorkomt, dan maakt het voor ideologie waarschijnlijk meer
293 uit of er *minder asfalt* of *meer asfalt* staat.

294 **Support Vector Machines en Logistische Regressie** De meest voorko-
295 mende techniek in vergelijkbaar onderzoek is Support Vector Machine (SVM).
296 Een andere techniek die gebruikt wordt is logistische regressie. Beide kennen
297 een eigen implementatie in sklearn, maar deze implementaties zijn niet efficiënt met
298 grote datasets. Om deze reden is er in beide gevallen voor gekozen om gebruik
299 te maken van de functie `SGDClassifier`, die beide technieken leert met *stochastic*
300 *gradient descent learning*. Er is hiervoor gevarieerd met de regularisatie, learning
301 rate en maximum aantal iteraties. Voor regularisatie is hier geëxperimenteerd
302 met Lasso en Ridge regularisatie, en een combinatie van beide genaamd Elastic-
303 net. De andere parameters zijn gelaten op de standaardwaarden van scikit-learn
304 [9].

305 **Naive Bayes** Een simpelere techniek die gebruikt wordt voor politieke tekst-
 306 classificatie is Naive Bayes. Dit algoritme neemt aan dat elke *feature* onafhan-
 307 kelijk is ten op zichte van de rest. Dit is bij tekstclassificatie vaak niet het geval
 308 omdat het gebruik van sommige woorden gepaard kan gaan met het gebruik
 309 van andere woorden. Daarnaast is het gebruik van meerdere n-grams in een
 310 classificatie schending van de aanname, want als bijvoorbeeld een bigram er in
 311 voorkomt dan komen ook beide unigrams er sowieso in voor. Desalniettemin
 312 blijkt Naive Bayes effectief te zijn voor tekstclassificatie[9, 6]. Hiervoor zijn de
 313 functies van scikit-learn MultinomialNB en BernoulliNB gebruikt.[9, 6]

314 **Beoordelen van kwaliteit** De meest gebruikte methoden om kwaliteit van
 315 politieke tekstclassificatie te beoordelen zijn *accuracy* en F_1 score, die opge-
 316 bouwd is uit recall en precision. Deze scores zijn opgebouwd uit vier variabelen.
 317 Deze variabelen geven weer hoeveel documenten wel of niet bij een klasse horen,
 318 en of deze wel of niet als dusdanig zijn geclassificeerd [10] .

	Behorend tot partij	Niet behorend tot partij
Geclassificeerd als partij	<i>true positive (tp)</i>	<i>false positive (fp)</i>
Niet geclassificeerd als partij	<i>false negative (fn)</i>	<i>true negative (tn)</i>

$$Precision = \frac{tp}{tp + fp} \quad (1)$$

$$Recall = \frac{tp}{tp + tn} \quad (2)$$

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad (3)$$

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

322 *Accuracy* is het percentage van documenten dat correct geclassificeerd is. Preci-
 323 sion is het percentage van documenten geclassificeerd als klasse, dat ook bij die
 324 klasse hoort. Recall is het percentage documenten van documenten behorende
 325 tot een klasse, dat ook als dusdanig geclassificeerd is. F_1 is het harmonisch
 326 gemiddelde van recall en precision. Precision, recall en dus ook F_1 worden per
 327 klasse berekend. Er zijn drie varianten om deze scores voor de hele classificatie
 328 te berekenen.

329 Allereerst is er *micro*, daarbij worden alle waarden bij elkaar opgeteld en
 330 dan berekend. Dit leidt ertoe dat resultaten van klassen met veel documenten
 331 belangrijker zijn. Als een classificatie kleine klassen grotendeels fout classificeert,
 332 kan deze score alsnog hoog zijn. In het geval van meer dan twee klassen is dit
 333 hetzelfde als *accuracy*.

334 Als tweede is er *macro*, daarbij worden alle scores per klasse berekend en
 335 wordt daarvan het gemiddelde genomen. Dit leidt er dan weer toe dat resultaten
 336 van klassen met weinig documenten net zo belangrijk zijn. Hierdoor kan een
 337 classificatie met een laag aantal correct geclassificeerde documenten hoog scoren
 338 door vooral kleine klassen goed te classificeren.

339 Als laatste is er dan nog *gewogen*, deze berekent net als *macro* de scores per
 340 klasse, maar neemt hiervan het gemiddelde gewogen bij het aantal documenten

341 behorend tot een klasse. Deze wijkt weinig af van de *micro* variant, tenzij er
342 uitschieters zijn bij klassen.

343 Aangezien *micro* al terugkomt in *accuracy* en het nadeel van *macro* te
344 groot is omdat de klassen nogal variëren in grootte, is gekozen voor *gewogen F_1*
345 scoring naast *accuracy*.

346 3.2.2 Deelvraag 2

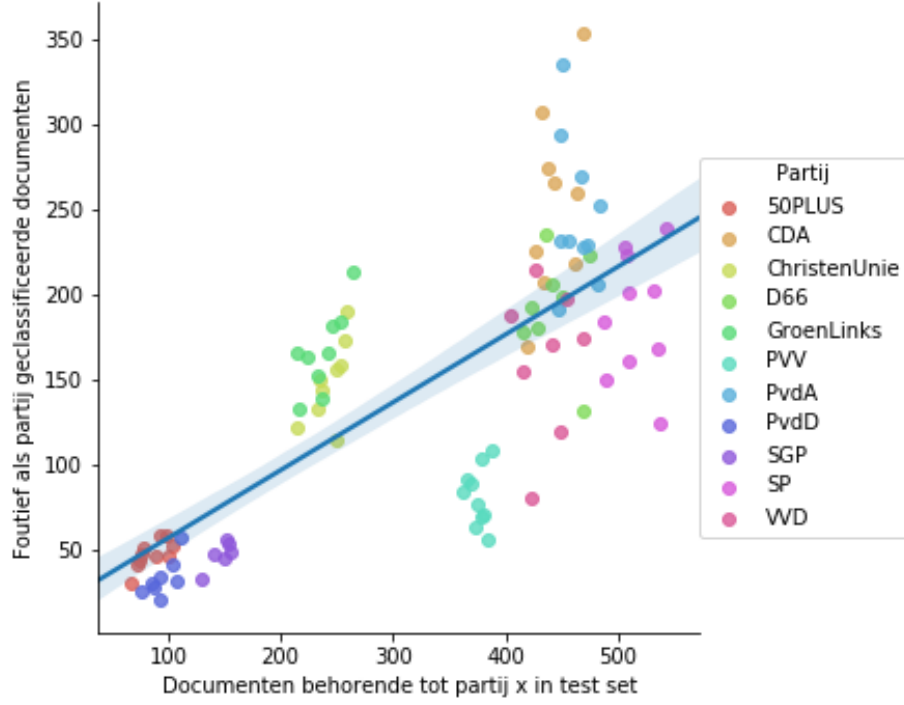
347 In Diermeier et al. [4] wordt aangenomen dat namen een groot effect hebben op
348 de classificatie en Hirst et al. [2] bevestigen dit voor het Europees Parlement.
349 Aangezien hier bij deelvraag 1 niet voor is gekozen, wordt bij deze deelvraag
350 gekeken hoe groot het effect hiervan is, specifiek gericht op partijnamen en ach-
351 ternamen van Kamerleden. Voor deze deelvraag wordt wederom een classificatie
352 gedaan met de classificatiemethode die resulteerde uit deelvraag 1. In deze clas-
353 sificatie worden alle partijnamen vervangen door de tag PARTIJNAAM en alle
354 namen van Kamerleden vervangen door de KAMERLIDNAAM. Deze namen
355 zijn uit de Handelingen gehaald. Voor partijnamen zijn ook lidwoorden toege-
356 voegd, voor achternamen van Kamerleden zijn ook verkortingen meegenomen.
357 Dit laatste omdat bijvoorbeeld *Van Haersma Buma* vaak aangesproken wordt
358 als *Buma*. Voornamen van Kamerleden worden zelden tot nooit gebruikt, dus
359 die zijn er niet uitgehaald. Een nadeel van deze aanpak is dat ook namen van
360 niet-Kamerleden of andere woorden weggehaald kunnen worden als deze het-
361 zelfde zijn als naam van een Kamerlid. Door gebruik van gevoeligheid voor
362 hoofdletters is geprobeerd dit te voorkomen. Een opvallend voorbeeld hiervan
363 is de naam Rutte, die zowel behoort tot het Kamerlid Arno Rutte als de premier
364 Mark Rutte. Steekproefgewijs is gekeken of er nog namen achter zijn gebleven,
365 maar die zijn niet gevonden.

366 Ook wordt gekeken naar classificatie met alleen partijnamen en namen van
367 Kamerleden. Alle andere woorden worden weggehaald. Namen van Kamerleden
368 en partijen die niet aan elkaar geschreven worden, zoals *Partij van de Arbeid*,
369 worden aan elkaar geschreven zodat het één feature wordt. Doordat alle andere
370 woorden weggehaald zijn, worden de bi- en trigrams combinaties van namen
371 die zinnen uit elkaar kunnen staan, dus die niet meer informatie geven dan
372 unigrams. Daarom wordt er gebruikt van de classificatiemethode uit deelvraag
373 1, maar dan met alleen unigrams. Hoge scores voor deze classificatie geven
374 aan dat met alleen namen classificatie goed te doen is en dat dit dus een grote
375 bijdrage heeft geleverd aan de resultaten uit deelvraag 1.

376 3.2.3 Deelvraag 3

377 Om deze deelvraag te beantwoorden zal een analyse gedaan worden van de
378 confusion matrix en zullen twee experimenten die Graeme Hirst et al. uitvoerden
379 voor dezelfde vraag gereproduceerd worden op de dataset van de Tweede Kamer.
380 Bij deze deelvraag zal de beste classifier uit deelvraag 1 en 2 gebruikt worden.

381 Als er een confounding bias is op basis van partij-status, dan is te verwach-
382 ten dat het aantal misclassificaties minus verwachte waarde binnen regerings-
383 partijen en binnen oppositiepartijen hoger ligt dan tussen oppositiepartijen en
384 regeringspartijen. Uit de voorverkenning (op basis van resultaten uit deelvraag
385 1 en 2) blijkt verder dat er een correlatie is tussen het aantal *false positives* van
386 een partij en het aantal documenten behorend tot die partij.



Figuur 2: Het aantal foutief als bepaalde partij geclassificeerde documenten ten opzichte van het aantal documenten behorend tot die partij. Dit is op basis van 50 classificaties met verschillende test en train set. De pearson correlatie is 0.78.

Op basis van dit verband is het verwachte aantal documenten

$$V_{i,j} = fn_i * \frac{tp_j + fn_j}{tn_i + fp_i} \quad (5)$$

waar $i \neq j$ met i de voorspelde partij en j de echte partij waar een document bijhoort.

De error is dan het verschil van de verwachte waarde en het daadwerkelijk aantal documenten

$$e_{i,j} = D_{i,j} - V_{i,j} \quad (6)$$

met opnieuw $i \neq j$ en i de voorspelde partij en j de echte partij waar een document bijhoort.

Als dit een goede benadering is van de error, dan is het te verwachten dat deze normaal verdeeld is [11]. Om te kijken of er een confounding bias is, worden de distributies binnen regeringspartijen, binnen oppositiepartijen en tussen beide groepen met elkaar vergeleken. Om de invloed van variantie door de willekeurige splitsing documenten voor trainen en testen te beperken, wordt de classificatie 50 keer gedaan en worden deze errors bij elkaar in distributie genomen. De nulhypothese is dat er geen verschil is tussen de verdelingen. De alternatieve hypothese is dan dus dat er wel een verschil is tussen de verdelingen. Als de nulhypothese wordt verworpen, kan dus aangenomen worden dat er een verschil is op basis van partij-status.

In het eerste experiment uit Graeme Hirst et al. zullen de tien meest karakteristieke woorden per partij van de ene zittingsperiode vergeleken worden met de tien meest karakteristieke woorden per partij van de andere zittingsperiode. Als de classificatie op basis van ideologie is in plaats van partij-status, is het te verwachten dat de woorden bij een partij blijven en niet gekoppeld zijn aan in oppositie of regering zitten.

In het tweede experiment uit Graeme Hirst et al. worden classifiers getraind op de ene zittingsperiode en getest op de andere zittingsperiode. Als de classificatie op basis van ideologie is in plaats van partij-status, is de verwachting dat er nog steeds aanzienlijke voorspellingen gedaan worden, aangezien de ideologie naar verwachting redelijk stabiel is binnen tien jaar (hoewel woordgebruik varieert). Als de scores aanzienlijk lager zijn, kan dit het gevolg zijn van het veranderen van partij-status van partijen.

Als vergelijkingsmateriaal is voor deze experimenten een tweede dataset nodig uit een ander kabinet. Hiervoor is het wenselijk dat dit kabinet bestaat uit andere partijen dan kabinet-Rutte II. Daarnaast is het ook wenselijk als het niet te ver terug is, zodat onderwerpen en taalgebruik enigszins overeenkomstig zijn. Omdat kabinet-Rutte I een minderheidskabinet was met een bijzondere partij-status voor de PVV, is ervoor gekozen om de Handelingen van de Tweede Kamer tijdens het missionaire kabinet-Balkenende IV (22 februari 2007 tot 20 februari 2010) te gebruiken.

De partij 50PLUS bestond nog niet gedurende kabinet-Balkenende IV, dus documenten van deze partij zijn weggelaten. Verder heeft dezelfde verwerking van data plaatsgevonden, zoals beschreven in 3.1. Alleen de minimum- en maximumlengte is overgenomen van de dataset van kabinet-Rutte II.

3.2.4 Deelvraag 4

Voor deze deelvraag vergelijken we de resultaten van de eerdere classificatie per partij met een binaire classificatie op basis van rechts en links. Hiervoor wordt wederom de dataset van kabinet-Rutte 2 gebruikt, met het beste model wat resulteerde uit deelvraag 1.

Voor deze vraag moet vastgesteld worden welke partijen links en rechts zijn. Omdat dit lastig te bepalen is en er meerdere indelingen zijn, wordt hier gebruik gemaakt van twee verschillende indelingen. De indeling op basis van het Kieskompas van Andre Krouwel voor de Kamerverkiezing 2012 en de indeling volgens het Manifesto Project[12] gebaseerd op verkiezingsprogramma's voor de Kamerverkiezing van 2012. In beide gevallen is de nullijn van het politieke spectrum gebruikt om te bepalen of een partij links of rechts is.

Tabel 2: Rechts (R) of link (L) indeling per partij op basis van het Kieskompas en het Manifesto Project.

Partij	Kieskompas	Manifesto Project
SP	L	L
PvdA	L	L
GroenLinks	L	L
PvdD	L	L
50PLUS	L	L
D66	R	L
PVV	-	R
ChristenUnie	R	R
SGP	R	R
VVD	R	R
CDA	R	R

4 Evaluatie

4.1 Resultaten

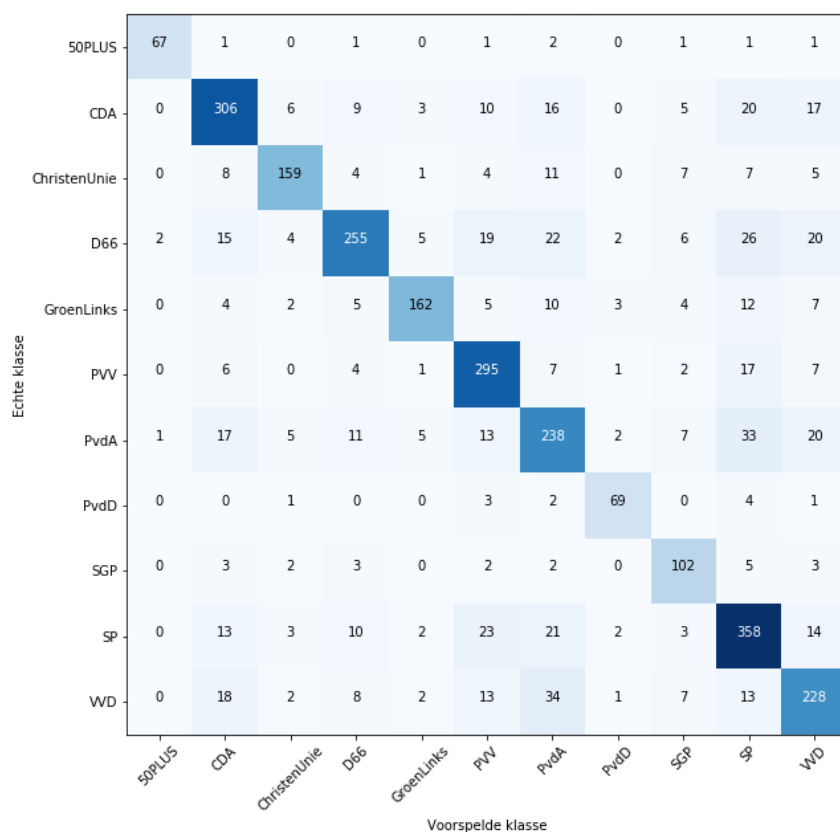
4.1.1 Deelvraag 1

Het beste resultaat werd bereikt met SVM gebruikmakend van *stochastic gradient descent learning* en Ridge regularisatie. De scores zijn ruim hoger dan de baseline scores. De scores liggen binnen de scores gevonden in gerelateerd werk, ondanks dat de baseline scores aanzienlijk lager zijn.

Tabel 3 laat de classificatiescores zien per partij met het aantal documenten in de test set. De F_1 scores per partij liggen tussen de 0.7 en 0.8. De one-issuepartijen, 50PLUS en PvdD, hebben scores daarboven, terwijl de coalitiepartijen, VVD en PvdA, lagere scores hebben. Figuur 3 laat zien waar de fouten in deze classificatie zitten.

Tabel 3: Classificatierapport van beste classificatie.

Partij	Precision	Recall	F1_score	Documenten
50PLUS	0.908	0.868	0.886	77.8
CDA	0.780	0.774	0.776	396.4
ChristenUnie	0.852	0.756	0.798	211.4
D66	0.814	0.672	0.734	380.0
GroenLinks	0.880	0.744	0.804	218.6
PVV	0.754	0.856	0.802	345.6
PvdA	0.652	0.670	0.658	354.4
PvdD	0.826	0.834	0.830	83.4
SGP	0.702	0.812	0.748	126.2
SP	0.724	0.786	0.748	456.2
VVD	0.700	0.692	0.696	330.0



Figuur 3: Confusion matrix van beste classificatie.

453

De meest karakteristieke features per partij zijn te zien in tabel 4.

Tabel 4: Meest relevante woorden per partij op basis van beste classificatie gedurende kabinet-Rutte II.

50PLUS	CDA	ChristenUnie	D66	GroenLinks
50plus	cda	de christenunie	d66	groenlink
lid krol	het cda	christenunie	led van veldhov	lid van tonger
het lid krol	cda fractie	het lid dik	lid van veldhov	het lid voortman
lid krol nar	de cda fractie	lid dik	lid van men	lid voortman
krol nar mij	de cda	lid dik faber	d66 is	lid voortman nar
krol nar	lid omtzigt	de led dik	led bergkamp	tonger nar
van 50plus	het lid omtzigt	led dik faber	de led bergkamp	tonger nar mij
gepensioneerd	lid omtzigt nar	led dik	led van men	van tonger nar
krol	het cda is	de led voordewind	d66 wil	led van tonger
50plus is	cda is	led voordewind	d66 vindt	de led voortman

Tabel 4: Meest relevante woorden per partij op basis van beste classificatie gedurende kabinet-Rutte II. (*Vervolg*)

PVV	PvdA	PvdD	SGP	SP	VVD
pvv	de pvda	lid ouwehand nar	de sgp	sp	de vvd
de pvv	pvda	lid ouwehand	sgp	de sp	vvd
islamitisch	de partij van	het lid ouwehand	sgp fractie	lid van gerv	de vvd is
miljard	van de arbeid	ouwehand nar	de sgp fractie	sp fractie	vvd is
lid graus	pvda fractie	ouwehand nar mij	de led dijkgraf	gerv nar mij	de vvd fractie
het lid graus	de arbeid	ouwehand	led dijkgraf	van gerv nar	vvd fractie
lid graus nar	de pvda fractie	vor de dier	led van der	gerv nar	vor de vvd
madlener nar mij	partij van de	de dier	led bisschop	de sp fractie	wat de vvd
madlener nar	partij van	dier	de led bisschop	gerv	vvd vindt
het lid madlener	arbeid	thiem	de sgp is	van gerv	de vvd vindt

454 Hierin is te zien dat vrijwel alle woorden verwijzen naar de partij of een
 455 Kamerlid van die partij.

456 4.1.2 Deelvraag 2

457 In tabel 4 was al te zien dat de meest karakteristieke woorden voornamelijk
 458 bestaan uit partijnamen en namen van Kamerleden. In tabel 5 is vervolgens te
 459 zien welke woorden het meest karakteristiek zijn per partij, als partijnamen en
 460 namen van Kamerleden vervangen zijn.

Tabel 5: Meest relevante woorden per partij op basis van classificatie uit deelvraag 1 zonder partijnamen of namen van Kamerleden gedurende kabinet-Rutte II.

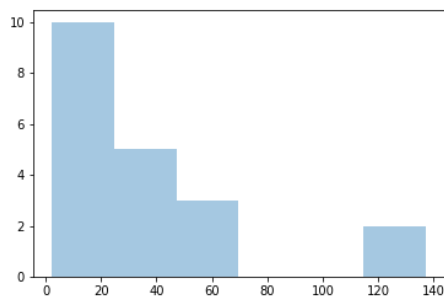
50PLUS	CDA	ChristenUnie	D66	GroenLinks
50 plusser	inwoner	voedselverspill	mijn fractie	schon energie
gepensioneerd	partijnam fractie	rookvrij	buitengewon	banenplan
plusser	onz inwoner	inderdad	daarom	schon
koopkrachtontwikkel	strak	gezinn	hervorm	huishoud zorg
exact	middeninkomen	loonwaard	natur	in elk geval
ouderenwerklos	hier	ik constater	kans	elk geval
50	bkk	als het gat	vind	belastingontwijk
de koopkrachtontwikkel	echt	geestelijk verzorger	scholing	kamer hierover te
werkend	partijnam	elkar	dadelijk	vergroen
ouder	diver	het ingezetenen criterium	was het voorzitter	zou

Tabel 5: Meest relevante woorden per partij op basis van classificatie uit deelvraag 1 zonder partijnamen of namen van Kamerleden gedurende kabinet-Rutte II. (*Vervolg*)

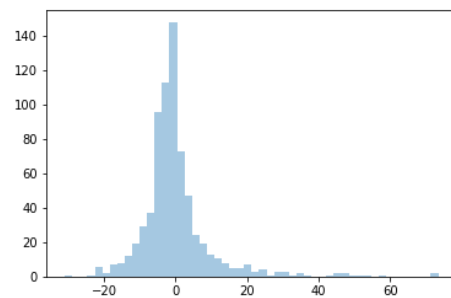
PVV	PvdA	PvdD	SGP	SP	VVD
islamitisch	ieder kind	dier	dank zer	segregatie	volgen mij
islam	mijn partij	de natur	mevrouw de voorzitter	huurder	liberal
de islam	circulair	bio industrie	mevrouw de	bureaucratie	essentieel
brussel	overigen	de bio industrie	eenverdiener	zegt	aangegev
asielzoeker	lager over	de bio	allerlei	de bevolk	kader
miljard	open standaard	bio	punt	herindel	speelveld
nederland	zegenschap	industrie	nadruk	geheim dienst	ondernemer
grenz	wijkverpleegkund	aan de bio	stevig	groeïend	ban
immigratie	tevred	dierenwelzijn	vanuit	bestuurder	partijnam fractie
dit kabinet	gezamen	moet kom aan	bepaald	zorginstell	haven

4.1.3 Deelvraag 3

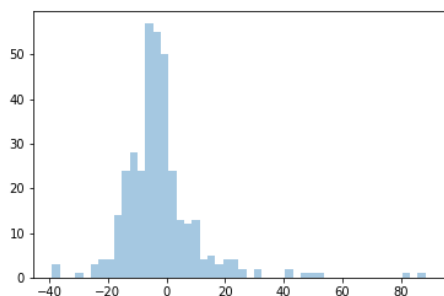
In figuur 4 zijn de distributies van de errors te zien van combinaties tussen regerings- en oppositiepartijen.



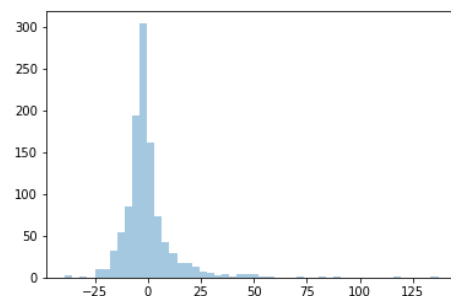
(a) Tussen twee regeringspartijen



(b) Tussen twee oppositiepartijen



(c) Tussen een regeringspartij en een oppositiepartij



(d) Totaal

Figuur 4: Distributie van de error uit 6 voor de verschillende combinaties.

464 4.2 Discussie

465 4.2.1 Deelvraag 1

466 Dit onderzoek heeft zich beperkt tot methoden genoemd in vergelijkbare onder-
467 zoeken en waarvan de implementatie beschikbaar is in Python. Een aantal me-
468 thoden die in gerelateerde literatuur leidden tot goede classificaties zijn daarom
469 niet getest. Ook nieuwe methoden die nog niet gebruikt zijn in een vergelijkbaar
470 onderzoek voor politieke tekst classificatie zijn daarom niet getest. Voor ver-
471 volgonderzoek kan daarom gekeken worden naar meer verschillende methoden.

472 Het onderzoek van Hirst et al. vond dat resultaten afhankelijk kunnen
473 zijn van documentgrootte. Alle documenten in dit onderzoek zijn kleiner dan
474 de grootste documentgrootte uit het onderzoek van Hirst et al. en ook de mi-
475 nimumfrequentie lager ligt dan de kleinste documentgrootte uit dat onderzoek.
476 Het effect wat zij vinden tussen documentgrootte van 267 en 6666 is een verschil
477 in nauwkeurigheid van 19,8%. Voor een vervolgonderzoek kan gekeken worden
478 naar of dit effect er is en wat dit betekent voor de resultaten.

479 4.2.2 Deelvraag 2

480 De resultaten laten zien dat de classificatie afhankelijk is van partijnamen en
481 namen van Kamerleden.

482 De woorden in tabel 5 komen bij veel partijen overeen met hun ideologie,
483 vooral bij PVV, PvdD en 50PLUS. Daarnaast zijn er ook woorden die niet
484 veel over ideologie zeggen, zoals; *volgens mij*, *ik constateer* en *in elk geval*.
485 Vooral de SGP heeft woorden die niet veel lijken te zeggen over de ideologie.
486 Met name opvallend hierbij is *mevrouw de voorzitter*, aangezien deze woorden
487 door alle partijen gebruikt worden om via de voorzitter te praten. Voor een
488 vervolgonderzoek kan gekeken naar waarom deze woorden zo karakteristiek zijn
489 voor partijen. Een hypothese is dat deze woorden eigen zijn aan een individueel
490 Kamerlid.

491 De classificatiemethode die gebruikt is in deze deelvraag, is gebaseerd op
492 de beste methode voor de dataset uit deelvraag 1. Hierin was gevonden dat een
493 combinatie van uni-, bi- en trigrams het beste resultaat opleverde. In tabel 4 is
494 te zien dat trigrams behoren tot de meest karakteristieke woorden, hoewel de
495 woorden in trigrams vaak overlappen met uni- en bigrams. In tabel 5 daaren-
496 tegen zijn er nog maar een paar trigrams, welke grotendeels procedurele zinnen
497 zijn of toevoeging van een lidwoord op een uni- of bigram. Dit verschil sugge-
498 reert dat trigrams minder belangrijk zijn in de classificatie zonder de namen,
499 dus de classificatiemethode uit deelvraag 1 niet het beste is voor deze classifica-
500 tie. In vervolgonderzoek kan de opzet van deelvraag 1 toegepast worden op de
501 classificatie zonder de namen, om zo te komen tot een classificatiemethode die
502 het beste resultaat oplevert op de classificatie zonder namen.

503 4.2.3 Deelvraag 4

504 Er zijn verschillende visies op links en rechts, en de indeling van de partijen,
505 ook buiten de twee methoden gekozen in dit onderzoek.

5 Conclusions

Referenties

- [1] Felix Bießmann. Automating political bias prediction. *CoRR*, abs/1608.02195, 2016.
- [2] Graeme Hirst, Yaroslav Riabinin, Jory Graham, and Magali Boizot-roche. Text to ideology or text to party status? *.
- [3] Vasco Ferreira. Using textual transcripts of parliamentary interventions for profiling portuguese politicians. 2016.
- [4] Daniel Diermeier, Jean-François Godbout, Bei Yu, and Stefan Kaufmann. Language and ideology in congress. *British Journal of Political Science*, 42(1):31–55, 2012.
- [5] Bei Yu, Stefan Kaufmann, and Daniel Diermeier. Classifying party affiliation from political speech. *Journal of Information Technology & Politics*, 5(1):33–48, 2008.
- [6] Conal Sathi Maneesh Bhand, Dan Robinson. Text classifiers for political ideologies, 2009.
- [7] Bjørn Høyland, Jean-François Godbout, Emanuele Lapponi, and Erik Vell-dal. Predicting party affiliations from european parliament debates. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 56–60. Association for Computational Linguistics, 2014.
- [8] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed ;today;].
- [9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [10] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.
- [11] *NIST/SEMATECH e-Handbook of Statistical Methods*. NIST/SEMATECH, April 2012.
- [12] Andrea Volkens, Pola Lehmann, Theres Matthieß, Nicolas Merz, Sven Regel, and Bernhard Weßels. The manifesto data collection. manifesto project (mrg/cmp/marpor). version 2017b, 2017.

A Slides