

1 IDEOLOGIE EN CLASSIFICATIE IN DE HANDELINGEN
2 VAN DE TWEEDE KAMER

3 SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF
4 BACHELOR OF SCIENCE

5 JASPER VAN DER HEIDE
6 10732721

7 BACHELOR INFORMATIEKUNDE
8 FACULTEIT DER NATUURWETENSCHAPPEN, WISKUNDE EN
9 INFORMATICA
10 UNIVERSITEIT VAN AMSTERDAM

11 2018-06-28

12

	Begeleider	Tweede lezer
Titel, Naam	Dr Maarten Marx	Ir Loek Stolwijk
Affiliatie	UvA, FNWI, IvI	UvA, FNWI, IvI
Email	M.J.Marx@uva.nl	A.M.Stolwijk@uva.nl



UNIVERSITEIT VAN AMSTERDAM

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	8
24	3.2.1 DV1: Beste classificatiemethode	8
25	3.2.2 DV2: Invloed van namen	10
26	3.2.3 DV3: Oppositie of regering	11
27	3.2.4 DV4: Links of rechts	13
28	3.2.5 DV5: Woordgebruik van sprekers	14
29	4 Evaluatie	14
30	4.1 Resultaten	14
31	4.1.1 DV1: Beste classificatiemethode	14
32	4.1.2 DV2: Invloed van namen	17
33	4.1.3 DV3: Oppositie of regering	19
34	4.1.4 DV4: Links of rechts	21
35	4.1.5 DV5: Woordgebruik van sprekers	21
36	4.2 Discussie	21
37	4.2.1 DV1: Beste classificatiemethode	21
38	4.2.2 DV2: Invloed van namen	22
39	4.2.3 DV3: Oppositie of regering	22
40	4.2.4 DV4: Links of rechts	23
41	4.2.5 DV5: Woordgebruik van sprekers	23
42	4.2.6 Algemeen	23
43	5 Conclusies	23
44	A Slides	24

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?
5. In hoeverre wordt deze classificatie bepaald door woordgebruik van sprekers?

Voor de eerste deelvraag zullen Support Vector Machine, Logistische Regressie en Naive Bayes vergeleken worden aan de hand van *accuracy* en F_1 score. Bij de tweede deelvraag wordt gekeken naar het effect van het weglaten van partijnamen en namen van Kamerleden. De derde vraag bestaat uit meerdere experimenten, waarin gekeken zal worden naar of de misclassificaties binnen coalitie of oppositie groter zijn dan daartussen, en of er tussen die groepen verschillen zitten in de confusion matrix.

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.

92 2 Gerelateerd werk

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

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

102 2.1 Tekstclassificatie van parlementaire teksten

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

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

116 Yu et al. [5] richtte zich vervolgens op zowel het Amerikaanse Huis van
117 Afgevaardigden als de Senaat in 2005. Een document was in dit onderzoek de
118 verzameling van alle speeches van een senator in een Congres en het label de
119 partij. Voor het Huis van Afgevaardigden vonden ze een nauwkeurigheid van
120 80.1% (baseline van 51.5%) en voor de senaat 86.0 % (baseline van 55.0%). Ze
121 testten hun classificaties ook op de andere kamer. Van Huis van Afgevaardigden
122 naar senaat leverde dit een nauwkeurigheid op van 88.0% (baseline van 55.0%)
123 en andersom 67.6% (baseline van 51.5%). Hun verklaring voor dit verschil is
124 dat het Huis van Afgevaardigden meer partisan is.

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

131 Als een vervolg op deze onderzoeken deden Graeme Hirst et al. een verge-
132 lijikbaar onderzoek naar het Canadese Parlement [2]. Hierbij werd zowel gekeken
133 naar de Engelse als Franse teksten. Een document werd hier gezien als de samen-
134 voeging van alle spreekbeurten van een spreker. Afhankelijk van taal en dataset
135 vinden zij in dit onderzoek nauwkeurigheden van 83.2% en hoger (baseline van
136 65.5%).

137 Het onderzoek bevat ook een classificatie van het Europees Parlement.
138 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

Diermeier et al. [4] gebruikten 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*.

Yu et al. [5] maakten 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 wegen 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.

Graeme Hirst et al. maakten 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 wegen voor woorden: *boolean*, *tf*, *tf-norm* en *tf-idf*, waarvan *tf-idf* het beste resultaat opleverde.

Bhand et al. gebruikten 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 wegen van woorden werd

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

190 Høyland et al. maakten gebruik van Support Vector Machine[7]. Als beste
191 waarde voor de regularisatieterm, de C-parameter, vonden zij 0.8. Daarnaast
192 gebruikten zij *dependency disambiguated stems* wat bij hen een F_1 score van
193 twee procent hoger opleverden dan normale stemming.

194 2.3 Invloed van partijnamen of sprekersnamen

195 Diermeier et al. lieten de namen van de sprekers en verwijzingen naar staten
196 die de senatoren representeren weg, omdat deze volgens hen de classificatie te
197 makkelijk zouden maken [4]. Hirst et al. vinden inderdaad dat partijnamen (en
198 het weglaten daarvan) bij het Europees Parlement een grote invloed hebben op
199 de classificatie [2]. Bij het Europees Parlement zien zij met name het gebruik
200 van de eigen partijnaam door een spreker, terwijl zij in het Canadese parlement
201 vooral zien dat de naam van de andere partij gebruikt wordt door een spreker.

202 2.4 Invloed van oppositie of regering

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

212 In hetzelfde onderzoek trainden ze ook hun classifiers op het ene parle-
213 ment en testten deze op het andere parlement. Hierbij vonden zij in beide
214 gevallen een nauwkeurigheid ver onder de baseline. Daarnaast deden ze ook
215 nog een classificatie op de sprekers die in beide parlementen zaten en een an-
216 dere classificatie op sprekers die niet in beide parlementen zaten. Bij de eerste
217 classificatie vonden ze nauwkeurigheden rond de baseline, terwijl in de tweede
218 situatie nauwkeurigheden gevonden werden ver boven de baseline.

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

221 3 Methodologie

222 3.1 De data

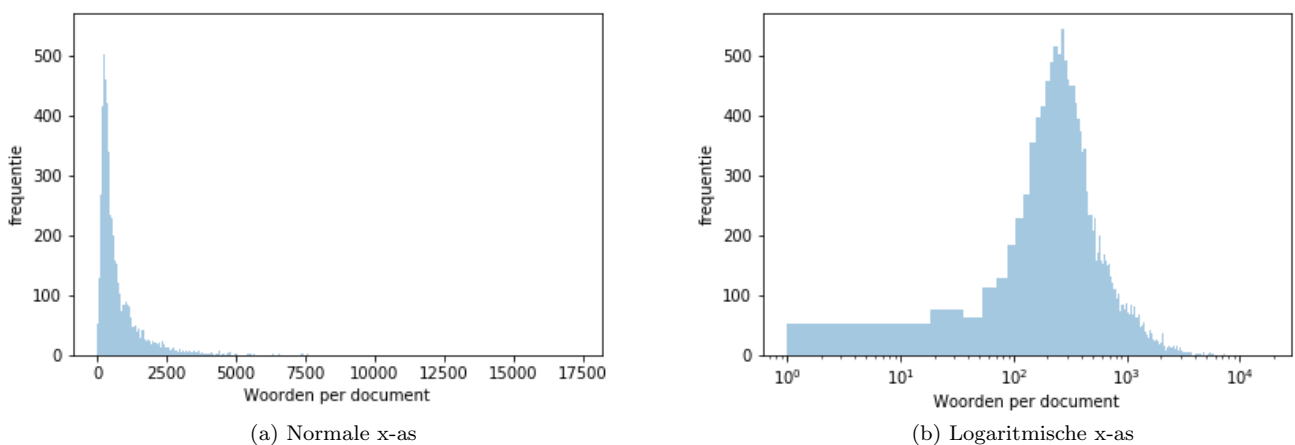
223 De data die gebruikt worden, zijn de Handelingen van de Tweede Kamer gedu-
224 rende het missionaire kabinet-Rutte II (5 november 2012 tot 22 maart 2017).
225 Er is gekozen voor dit kabinet, omdat de data hiervoor makkelijk verkrijgbaar
226 was, het kabinet lang zat - waardoor er veel data is - en het recent is waar-
227 door het makkelijker te interpreteren is. Deze data zijn in xml-formaat van de

website officiële bekendmakingen.nl gehaald samen met corresponderende meta-data xml-bestanden. De bestanden van de Handelingen bevatten voornamelijk informatie over spreekbeurten tijdens een debat, waaronder naam van een spreker, partij-affiliatie, inhoud van de spreekbeurt en het soort spreekbeurt. Deze gegevens zijn samengevoegd tot één tabel.

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

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

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



Figuur 1: Aantal woorden per document

Om toch de uitschieters er uit te halen, is aangenomen dat het wel lognor-

258 maal verdeeld is en zijn daarmee de documenten buiten het betrouwbaarheids-
 259 interval van 95% eruit gehaald. De documenten met een lengte van minimaal
 260 28 en maximaal 1492 woorden bleven daarmee over. Het gemiddelde is daarna
 261 498 woorden en de mediaan is 386 woorden. Een totaal aantal documenten van
 262 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

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

268 3.2 Methoden

269 3.2.1 DV1: Beste classificatiemethode

270 Om deze deelvraag te beantwoorden zullen een aantal classificatiemethoden ver-
 271 geleken worden. Aangezien het onmogelijk is om alle classificatiemethoden te
 272 vergelijken, beperkt dit onderzoek zich tot classificatiemethoden die gebruikt
 273 zijn in vergelijkbare onderzoeken, zoals besproken in 2.2. Er is ervoor geko-
 274 zen om alleen gebruik te maken van methoden waarvan reeds implementaties
 275 beschikbaar waren in scikit-learn. Voor alle methoden wordt gezocht naar de
 276 beste parameters; een grid search. Deze grid search wordt gedaan door middel
 277 van 5-fold cross-validation, waarbij de trainings set steeds 80% is en de test set
 278 20% van de totale dataset. De hypothese is dat de scores lager zijn dan die
 279 gevonden in het gerelateerd werk, omdat de documentgrootte kleiner is en de
 280 baseline lager.

281 **Pre-processing** Voor pre-processing is gebruik gemaakt van tokenisation en
 282 lowercasing. Voor tokenisation is de reguliere expressie
 283 $w+$ gebruikt, waardoor alles behalve letters en cijfers weggehaald wordt. Ver-
 284 volgens is er gevarieerd tussen wel of geen gebruik maken van stemming. In

285 het geval van stemming is gebruik gemaakt van de Snowball Stemmer via de
286 Python NLTK module.

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

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

303 **Support Vector Machines en Logistische Regressie** De meest voorko-
304 mende techniek in vergelijkbaar onderzoek is Support Vector Machine (SVM).
305 Een andere techniek die gebruikt wordt is logistische regressie. Beide kennen een
306 eigen implementatie in scikit-learn, maar deze implementaties zijn niet efficiënt
307 met grote datasets. Om deze reden is er in beide gevallen voor gekozen om
308 gebruik te maken van de functie *SGDClassifier*, die beide technieken leert met
309 *stochastic gradient descent learning*. Voor regularisatie is hier geëxperimenteerd
310 met Lasso en Ridge regularisatie, en een combinatie van beide genaamd Elastic-
311 net. De andere parameters zijn gelaten op de standaardwaarden van scikit-learn
312 [9]. Een belangrijke onaangepaste waarde is die van maximaal aantal iteraties,
313 die als standaard 5 heeft. Volgens scikit-learn convergeert de *SGDClassifier*
314 rond de $10^6/n$ iteraties waar n het aantal documenten in de training set is. In
315 het geval van deze dataset zou dat 84 iteraties zijn. Vanwege de grootte van de
316 gridsearch was het voor dit onderzoek niet mogelijk het maximum iteraties te
317 verhogen.

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

327 **Beoordelen van kwaliteit** De meest gebruikte methoden om kwaliteit van
328 politieke tekstclassificatie te beoordelen zijn *accuracy* en F_1 score, die opge-
329 bouwd is uit recall en precision. Deze scores zijn opgebouwd uit vier variabelen.

330 Deze variabelen geven weer hoeveel documenten wel of niet bij een klasse horen,
 331 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>

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

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

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

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

335 *Accuracy* is het percentage van documenten dat correct geclassificeerd is. *Precision*
 336 is het percentage van documenten geclassificeerd als klasse, dat ook bij die
 337 klasse hoort. *Recall* is het percentage documenten van documenten behorende
 338 tot een klasse, dat ook als dusdanig geclassificeerd is. F_1 is het harmonisch
 339 gemiddelde van recall en precision. Precision, recall en dus ook F_1 worden per
 340 klasse berekend. Er zijn drie varianten om deze scores voor de hele classificatie
 341 te berekenen.

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

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

352 Als laatste is er dan nog *gewogen*, deze berekent net als *macro* de scores per
 353 klasse, maar neemt hiervan het gemiddelde gewogen bij het aantal documenten
 354 behorend tot een klasse. Deze wijkt weinig af van de *micro* variant, tenzij er
 355 uitschieters zijn bij klassen.

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

359 3.2.2 DV2: Invloed van namen

360 In Diermeier et al. [4] wordt aangenomen dat namen een groot effect hebben op
 361 de classificatie en Hirst et al. [2] bevestigen dit voor het Europees Parlement.
 362 Aangezien hier bij deelvraag 1 niet voor is gekozen, wordt bij deze deelvraag
 363 gekeken hoe groot het effect hiervan is, specifiek gericht op partijnamen en ach-
 364 ternamen van Kamerleden. Voor deze deelvraag wordt wederom een classificatie

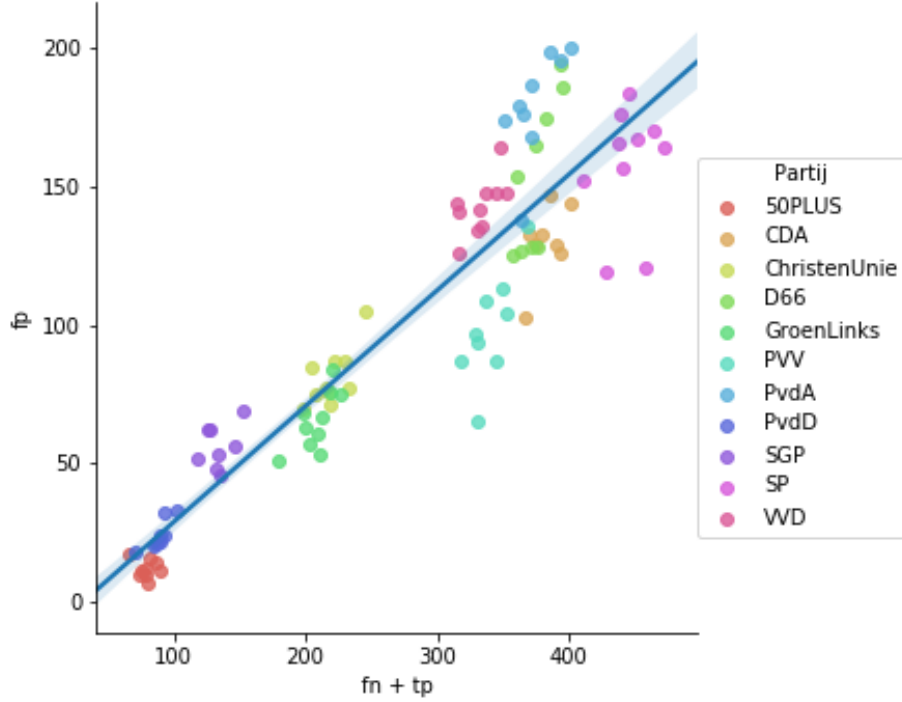
gedaan met de classificatiemethode die resulteerde uit deelvraag 1. In deze classificatie worden alle partijnamen vervangen door de tag PARTIJNAAM en alle namen van Kamerleden vervangen door de KAMERLIDNAAM. Deze namen zijn uit de Handelingen gehaald. Voor partijnamen zijn ook lidwoorden toegevoegd, voor achternamen van Kamerleden zijn ook verkortingen meegenomen. Dit laatste omdat bijvoorbeeld *Van Haersma Buma* vaak aangesproken wordt als *Buma*. Voornamen van Kamerleden worden zelden tot nooit gebruikt, dus die zijn er niet uitgehaald. Een nadeel van deze aanpak is dat ook namen van niet-Kamerleden of andere woorden weggehaald kunnen worden als deze hetzelfde zijn als naam van een Kamerlid. Door gebruik van gevoeligheid voor hoofdletters is geprobeerd dit te voorkomen. Een opvallend voorbeeld hiervan is de naam Rutte, die zowel behoort tot het Kamerlid Arno Rutte als de premier Mark Rutte. Steekproefgewijs is gekeken of er nog namen achter zijn gebleven, maar die zijn niet gevonden.

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

3.2.3 DV3: Oppositie of regering

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

Als er een confounding bias is op basis van partij-status, dan is te verwachten dat het aantal misclassificaties minus verwachte waarde binnen regeringspartijen en binnen oppositiepartijen hoger ligt dan tussen oppositiepartijen en regeringspartijen. De verwachte waarde is afhankelijk van het aantal documenten van een partij in de training set [11]. Aangezien de test set uit dezelfde set als de training is gehaald, is de verwachte waarde ook afhankelijk van het aantal documenten van een partij in de test set. Uit de voorverkenning (op basis van resultaten uit deelvraag 1 en 2) blijkt deze correlatie tussen het aantal *false positives* van een partij en het aantal documenten behorend tot die partij.



Figuur 2: Het aantal *false positives* ten opzichte van het aantal documenten behorend tot die partij (*false negatives* en *true positives*). 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 j de voorspelde partij en i 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 [12]. 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 meest karakteristieke woorden per partij van de ene zittingsperiode vergeleken worden met de 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 afhankelijk is van partij-status is de verwachting dat de scores van partijen die gewisseld zijn van oppositie naar regering of andersom lagere scores krijgen dan partijen die niet van partij-status zijn veranderd.

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 DV4: Links of rechts

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[13] 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

3.2.5 DV5: Woordgebruik van sprekers

De vorige classificaties trainden op documenten en werden getest op andere documenten, maar wel van dezelfde sprekers als uit de training set. Naast de ideologie kan de classificatie daarom ook getraind zijn op het taalgebruik van sprekers. Als een Kamerlid bijvoorbeeld een woord regelmatig in speeches gebruikt, maar niet wordt gebruikt door zijn partijgenoten, wordt dit wel gezien als een belangrijk woord voor de partijclassificatie. Graeme Hirst et al. [2] plaatsten al een soortgelijke kanttekening bij de resultaten van Deiermeier et al.

Om te kijken of dit effect er is, wordt er opnieuw een classificatie gedaan. Ditmaal worden alleen niet de individuele documenten verdeeld over de training en test set, maar worden de Kamerleden, met bijbehorende documenten, verdeeld over de training en test set. Als taalgebruik van een spreker in de training set voorheen invloed had op de classificatie, zal dat nu geen effect meer hebben omdat er geen documenten van die spreker meer voorkomen in de test set. De meest karakteristieke woorden uit de resultaten van deelvraag 2 suggereren dat woordgebruik van Kamerleden invloed heeft (zie tabel 4). De hypothese is daarom ook dat deze nieuwe classificatie lagere scores vindt.

4 Evaluatie

4.1 Resultaten

4.1.1 DV1: Beste classificatiemethode

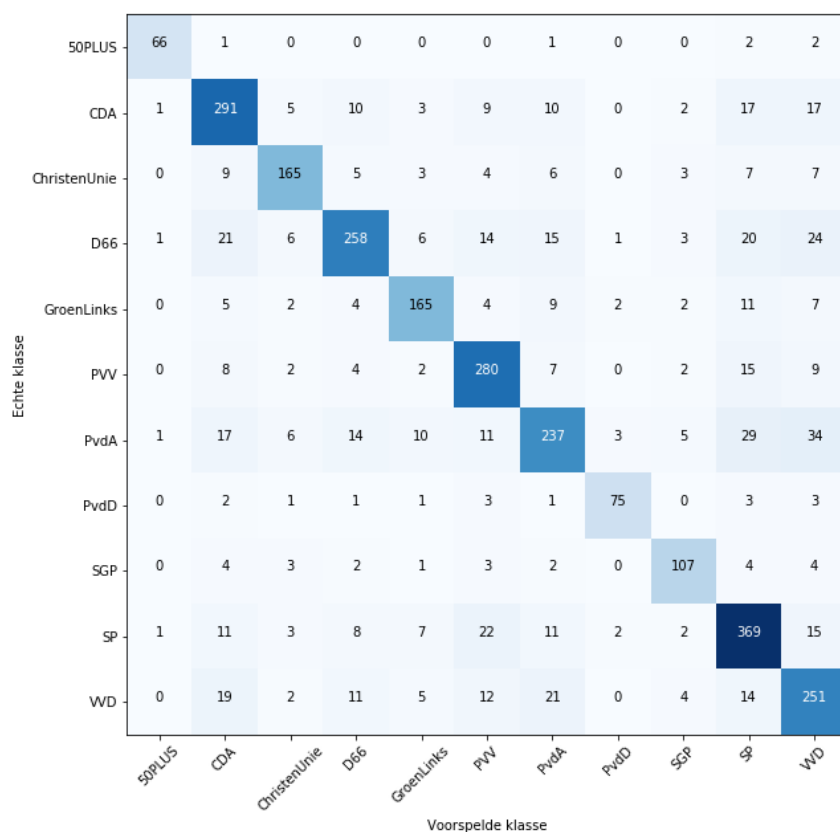
Het beste resultaat werd bereikt met Support Vector Machines gebruikmakend van *stochastic gradient descent learning* en Elasticnet regularisatie. De features waren hierbij gestemd, met unigrams, bigrams en trigrams. Geen features zijn hierin weggelaten door minimale of maximale documentfrequenties. 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 en de documentgrootte kleiner is.

Tabel 3 laat de scores 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,

484 50PLUS en PvdD, hebben scores daarboven, terwijl de coalitiepartijen, VVD
 485 en PvdA, lagere scores hebben. Figuur 3 laat zien waar de fouten in deze
 486 classificatie zitten. De meest karakteristieke features per partij zijn te zien in
 487 tabel 4. Hierin is te zien dat vrijwel alle woorden verwijzen naar de partij of
 488 een Kamerlid van die partij.

Tabel 3: Classificatierapport van beste classificatiemethode. Gemiddelde van vijf iteraties.

Partij	Precision	Recall	F1_score	Documenten
50PLUS	0.906	0.880	0.896	75.6
CDA	0.746	0.790	0.766	369.6
ChristenUnie	0.828	0.778	0.802	213.4
D66	0.806	0.694	0.746	373.4
GroenLinks	0.802	0.766	0.784	216.4
PVV	0.766	0.840	0.804	333.2
PvdA	0.738	0.638	0.682	372.2
PvdD	0.852	0.810	0.830	93.6
SGP	0.802	0.804	0.802	133.0
SP	0.744	0.812	0.778	455.6
VVD	0.668	0.732	0.698	344.0
avg / total	0.768	0.762	0.762	2980.0



Figuur 3: Confusion matrix van beste classificatie.

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 nar	het cda	christenunie	led van veldhov	lid van tonger
het lid krol	cda fractie	het lid dik	lid van men	led van tonger
lid krol	de cda fractie	lid dik	lid van veldhov	lid voortman
krol nar mij	de cda	lid dik faber	de led schouw	het lid voortman
krol nar	het cda is	led dik faber	led schouw	lid voortman nar
van 50plus	lid omtzigt nar	led dik	d66 is	tonger nar mij
fractie van 50plus	cda is	de led dik	d66 wil	tonger nar
krol	lid omtzigt	led voordewind	led van men	van tonger nar
50plus is	het lid omtzigt	de led voordewind	led bergkamp	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	sgp	sp	de vvd
de pvv	pvda	lid ouwehand nar	de sgp	de sp	vvd
islamitisch	de partij van	het lid ouwehand	sgp fractie	lid van gerv	de vvd fractie
lid graus	van de arbeid	ouwehand nar	de sgp fractie	sp fractie	vvd fractie
het lid graus	de arbeid	ouwehand nar mij	led dijkgraf	de sp fractie	de vvd is
lid graus nar	partij van de	ouwehand	de led dijkgraf	van gerv nar	vvd is
miljard	partij van	vor de dier	led van der	gerv nar	vor de vvd
graus nar mij	pvda fractie	de dier	de led bisschop	gerv nar mij	wat de vvd
graus nar	arbeid	dier	led bisschop	gerv	vvd betreft
madlener nar mij	de pvda fractie	de partij vor	sgp is	van gerv	de vvd betreft

4.1.2 DV2: Invloed van namen

In tabel 4 was al te zien dat de meest karakteristieke woorden voornamelijk bestaan uit partijnamen en namen van Kamerleden. In tabel 5 zijn de scores te zien van classificatie met partijnamen en namen van Kamerleden vervangen. Deze zijn aanzienlijk lager dan de scores uit deelvraag 1, maar nog steeds hoger dan de baseline. In tabel 6 is vervolgens te zien welke woorden het meest karakteristiek zijn per partij voor deze classificatie.

Tabel 5: Classificatierapport van beste classificatie zonder namen van Kamerleden of partijnamen.

	Precision	Recall	F1_score	Documenten
Partij				
50PLUS	0.658	0.454	0.524	76.2
CDA	0.492	0.418	0.446	370.8
ChristenUnie	0.516	0.322	0.384	218.8
D66	0.502	0.434	0.452	377.2
GroenLinks	0.486	0.338	0.368	210.8
PVV	0.512	0.768	0.612	349.8
PvdA	0.440	0.408	0.412	360.0
PvdD	0.554	0.596	0.572	89.0
SGP	0.592	0.660	0.620	134.8
SP	0.534	0.546	0.540	454.2
VVD	0.436	0.442	0.430	338.4
avg / total	0.502	0.486	0.480	2980.0

Tabel 6: Meest relevante woorden per partij op basis van classificatie uit deelvraag 1 zonder partijnamen of namen van Kamerleden gedurende kabinet-Rutte II.

CDA	ChristenUnie	D66	GroenLinks	PVV
PARTIJNAAM fractie	fractie van PARTIJNAAM	hervorm	PARTIJNAAMfractie	burger
wij	geweest	schon energie	deal	vreemdel
wij hebb	moment	dadelijk	politiek	belastingbetaler
buitengewon	de fractie van	arbeidsmarkt	klimaatverander	de burger
aangegev	volgen mij	de arbeidsmarkt	fractie van PARTIJNAAM	criminel
KAMERIIDNAAM	fractie van	outsider	je zou	belastinggeld
zorgvuld	de fractie	soort	klimaatcrisis	onz
dank	blij	plann	goed nieuw	gewon
zorgvrager	boeiend	generaties	enigszin	immigratie
via	PARTIJNAAMfractie	de patient	kilometerheff	natur

Tabel 6: Meest relevante woorden per partij op basis van classificatie uit deelvraag 1 zonder partijnamen of namen van Kamerleden gedurende kabinet-Rutte II. (Vervolg)

PvdA	PvdD	SGP	SP	VVD
vrouw	de bio	mijn fractie	student	onz fractie
medewerker	de bio industrie	diver	lerar	PARTIJNAAM
criteria	bio industrie	wel	voorstell	KAMERIIDNAAMman
roc	dier	vor hun beantwoord	ruimtelijk	essentieel
belangrijk	bio	helder	leerling	want
all	de veehouderij	uitdruk	mens	fractie zal
kinder	de natur	hun beantwoord	zegt	liberal
gezegd	landbouwareaal	allerlei	bureaucratie	PARTIJNAAM fractie
wij	crises	voorzitter ik wil	personeel	fractie
boet	dierenwelzijn	thema	ontslag	coalitie

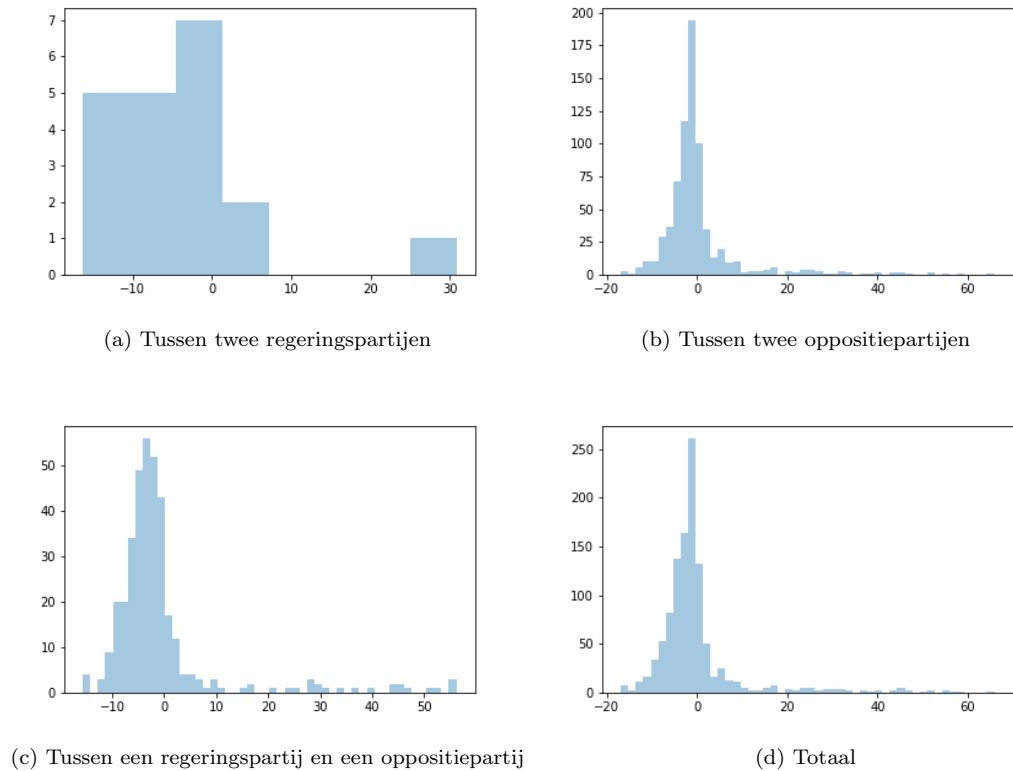
⁴⁹⁶ In tabel 7 zijn de scores te zien voor een classificatie met alleen namen van
⁴⁹⁷ partijen Kamerleden. De scores zijn gedaald ten op zichte van de resultaten van
⁴⁹⁸ deelvraag 1, maar hoger dan die zonder namen.

Tabel 7: Classificatierapport van beste classificatie met alleen namen van partijen en Kamerleden. Hiervoor is alleen gebruikgemaakt van unigrams.

Partij	Precision	Recall	F1_score	Documenten
50PLUS	0.822	0.832	0.828	73.0
CDA	0.680	0.650	0.664	370.0
ChristenUnie	0.638	0.594	0.614	215.4
D66	0.546	0.534	0.526	381.6
GroenLinks	0.630	0.662	0.626	208.2
PVV	0.670	0.700	0.670	346.6
PvdA	0.622	0.522	0.564	370.2
PvdD	0.766	0.796	0.782	88.2
SGP	0.730	0.464	0.568	133.2
SP	0.690	0.606	0.646	456.2
VVD	0.536	0.610	0.556	337.4
avg / total	0.642	0.610	0.616	2980.0

4.1.3 DV3: Oppositie of regering

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



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

Tabel 8: Meest relevante woorden per partij op basis van classificatie uit deelvraag 2 gedurende kabinet-Balkenende IV.

CDA	ChristenUnie	D66	GroenLinks	PVV
PARTIJNAAM fractie	fractie van PARTIJNAAM	hervorm	PARTIJNAAMfractie	burger
wij	geweest	schon energie	deal	de burger
buitengewon	de fractie van	ik hop	klimaatverander	immigratie
KAMERIIDNAAM	moment	welk	politiek	gewon
zorgvrager	fractie van	generaties	premier	vreemdel
aangegev	mijn fractie	patient	je zou	onz
helder	de fractie	outsider	fractie van PARTIJNAAM	land
PARTIJNAAM fractie heeft	uitgebreid	soort	mij	natur
zorgvuld	volgen mij	vor mij	de fractie	belastingbetal
wij hebb	blij	de arbeidsmarkt	tot slot	belastinggeld

Tabel 8: Meest relevante woorden per partij op basis van classificatie uit deelvraag 2 gedurende kabinet-Balkenende IV. (*Vervolg*)

PvdA	PvdD	SGP	SP	VVD
vrouw	dierproev	mijn fractie	lerar	onz fractie
belangrijk	bio industrie	dankzegg vor	informatie	PARTIJNAAM
roc	de bio industrie	dankzegg	personel	liberal
mijn collega	de bio	wel	zegt	KAMERIDNAAMman
hel duidelijk	bio	voorzitter ik wil	bureaucratie	fractie
groep	crises	uitdruk	ontslag	fractie zal
van de arbeid	veehouderij	positief	buurt	want
geweld	aan de bio	beantwoord	thuiszorg	PARTIJNAAM fractie
gezegd	eiwit	thema	bezuin	ondernemer
zer	de veehouderij	benader	ruimtelijk	markt

502 4.1.4 DV4: Links of rechts

503 4.1.5 DV5: Woordgebruik van sprekers

504 In tabel 9 staan de scores van classificatie waarbij de Kamerleden verdeeld zijn
505 over de training en test set. De scores zijn hierbij amper hoger dan de baseline.

Tabel 9: Classificatierapport van beste classificatie met de Kamerleden verdeeld over training en test set.

Partij	Precision	Recall	F1_score	Documenten
50PLUS	0.159	0.009	0.017	84.3
CDA	0.183	0.097	0.117	451.1
ChristenUnie	0.118	0.035	0.045	267.5
D66	0.214	0.286	0.202	328.8
GroenLinks	0.127	0.054	0.061	241.0
PVV	0.288	0.493	0.347	309.2
PvdA	0.179	0.143	0.155	354.2
PvdD	0.104	0.049	0.046	27.5
SGP	0.146	0.054	0.067	107.5
SP	0.286	0.289	0.277	522.1
VVD	0.215	0.280	0.232	328.9
avg / total	0.258	0.201	0.192	3022.1

506 4.2 Discussie

507 4.2.1 DV1: Beste classificatiemethode

508 Het onderzoek behaalt resultaten in lijn der verwachting op basis van gere-
509 lateerd en daarnaast ruim boven de baseline scores. De lage scores voor de
510 coalitiepartijen steunen de hypothese van een afhankelijkheid van partij-status,
511 zoals besproken wordt in deelvraag 3 Het bijna alleen voorkomen van partijna-
512 men en Kamerleden in de meest karakteristieke woorden per partij in tabel 4

513 steunt daarnaast het vermoeden dat deze classificatie sterk afhankelijk is van
514 die namen, zoals besproken wordt in deelvraag 2.

515 Dit onderzoek heeft zich beperkt tot methoden genoemd in vergelijkbare
516 onderzoeken en waarvan de implementatie beschikbaar is in scikit-learn. Een
517 aantal methoden die in gerelateerde literatuur leidden tot goede classificaties
518 zijn daarom niet getest. Ook nieuwe methoden die nog niet gebruikt zijn in een
519 vergelijkbaar onderzoek voor politieke tekst classificatie zijn daarom niet getest.
520 Daarnaast richtte zich dit ook maar op een beperkt aantal parameterwaarden.
521 Een belangrijke hierbij is het maximaal iteraties, wat ver onder het aantal ite-
522 raties benodigd voor convergentie ligt. Voor vervolgonderzoek kan daarom dit
523 onderdeel uitgebreid worden.

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

531 4.2.2 DV2: Invloed van namen

532 De resultaten laten zien dat de classificatie sterk afhankelijk is van partijna-
533 men en namen van Kamerleden. Deze daling was te verwachten op basis van
534 gerelateerd werk.

535 De woorden in tabel 6 komen bij veel partijen overeen met hun ideologie,
536 vooral bij PVV, PvdD en 50PLUS. Daarnaast zijn er ook woorden die niet veel
537 over ideologie zeggen, zoals; *volgens mij, ik constateer* en *in elk geval*. Vooral de
538 SGP heeft woorden die niet veel lijken te zeggen over de ideologie, hoewel deze
539 partij desalniettemin de hoogste f_1 score heeft. Met name opvallend hierbij
540 is *mevrouw de voorzitter*, aangezien deze woorden door alle partijen gebruikt
541 worden om via de voorzitter te praten. Voor een vervolgonderzoek kan gekeken
542 naar waarom deze woorden zo karakteristiek zijn voor partijen. Een hypothese
543 is dat deze woorden eigen zijn aan een individueel Kamerlid.

544 De classificatiemethode die gebruikt is in deze deelvraag, is gebaseerd op
545 de beste methode voor de dataset uit deelvraag 1. Hierin was gevonden dat een
546 combinatie van uni-, bi- en trigrams het beste resultaat opleverde. In tabel 4 is
547 te zien dat trigrams behoren tot de meest karakteristieke woorden, hoewel de
548 woorden in trigrams vaak overlappen met uni- en bigrams. In tabel 6 daaren-
549 tegen zijn er nog maar een paar trigrams, welke grotendeels procedurele zinnen
550 zijn of toevoeging van een lidwoord op een uni- of bigram. Dit verschil sugge-
551 reert dat trigrams minder belangrijk zijn in de classificatie zonder de namen,
552 dus de classificatiemethode uit deelvraag 1 niet het beste is voor deze classifica-
553 tie. In vervolgonderzoek kan de opzet van deelvraag 1 toegepast worden op de
554 classificatie zonder de namen, om zo te komen tot een classificatiemethode die
555 het beste resultaat oplevert op de classificatie zonder namen.

556 4.2.3 DV3: Oppositie of regering

557 In tabel 3 is het opvallend dat de coalitiepartijen lage scores krijgt. Daarnaast
558 laat figuur 3 zien dat er een hoge overlap zit tussen deze twee partijen.

559 4.2.4 DV4: Links of rechts

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

562 4.2.5 DV5: Woordgebruik van sprekers

563 De resultaten uit tabel 9 zijn laag, amper hoger dan de baseline. Dit suggereert
564 inderdaad dat eerdere classificaties in grote mate toch afhankelijk waren van het
565 woordgebruik van sprekers. Dit is opmerkelijk aangezien vergelijkbare werken
566 dit effect niet vinden.

567 Een alternatieve verklaring is dat de classificatie nu mede op basis van
568 woordvoerderschap is. Per onderwerp heeft een partij vaak maar één woord-
569 voerder, met uitzonderingen van wijzigingen in de fractie. Het is aannemelijk
570 dat het taalgebruik afhankelijk is van woordvoerderschap, aangezien er andere
571 termen gebruikt worden bij bijvoorbeeld een debat over zorg dan bij een debat
572 over onderwijs. Stel dat documenten van een spreker in de test set geclassifi-
573 ceerd moeten worden, dan kan het zijn dat deze meer karakteristieke vertoont
574 met een andere partij, aangezien er geen woordvoerder van die partij en dat
575 onderwerp in de training set zit, maar mogelijk wel van een andere partij. Een
576 vervolgonderzoek kan kijken of dit een verklaring is.

577 4.2.6 Algemeen

578 Het vergelijken van deze resultaten met vergelijkbaar werk is problematisch,
579 aangezien de keuzes en eigenschappen van hun onderzoek het niet een één-op-
580 één vergelijking maken. Voorbeelden hiervan zijn de documentgrootte, baseli-
581 nes, behouden of weglaten van namen, een spreker als document zien en het
582 trainen en testen op dezelfde spreker. Hoewel de resultaten dus lager zijn dan
583 die uit vergelijkbaar werk, moet hiermee rekening gehouden worden. Een ver-
584 volgonderzoek zou daarom dit onderzoek kunnen reproduceren op een ander
585 parlement om daarmee te kunnen vergelijken.

586 Dit onderzoek richtte zich hoofdzakelijk op de Handelingen gedurende
587 kabinet-Rutte II. Om te kijken in hoeverre het mogelijk is om deze conclu-
588 sie door te trekken naar de algemene Handelingen van de Tweede Kamer, kan
589 er in vervolgonderzoek gekeken worden naar meerdere zittingsperioden. Ook
590 kan gekeken worden naar veranderingen als een kabinet demissionair is.

591 5 Conclusies

592 Referenties

- 593 [1] Felix Bießmann. Automating political bias prediction. *CoRR*,
594 abs/1608.02195, 2016.
- 595 [2] Graeme Hirst, Yaroslav Riabinin, Jory Graham, and Magali Boizot-roche.
596 Text to ideology or text to party status? *.
- 597 [3] Vasco Ferreira. Using textual transcripts of parliamentary interventions for
598 profiling portuguese politicians. 2016.

- 599 [4] Daniel Diermeier, Jean-François Godbout, Bei Yu, and Stefan Kaufmann.
600 Language and ideology in congress. *British Journal of Political Science*,
601 42(1):31–55, 2012.
- 602 [5] Bei Yu, Stefan Kaufmann, and Daniel Diermeier. Classifying party affilia-
603 tion from political speech. *Journal of Information Technology & Politics*,
604 5(1):33–48, 2008.
- 605 [6] Conal Sathi Maneesh Bhand, Dan Robinson. Text classifiers for political
606 ideologies, 2009.
- 607 [7] Bjørn Høyland, Jean-François Godbout, Emanuele Lapponi, and Erik Vell-
608 dal. Predicting party affiliations from european parliament debates. In
609 *Proceedings of the ACL 2014 Workshop on Language Technologies and*
610 *Computational Social Science*, pages 56–60. Association for Computatio-
611 nal Linguistics, 2014.
- 612 [8] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source
613 scientific tools for Python, 2001–. [Online; accessed `today`].
- 614 [9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Gri-
615 sel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas,
616 A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay.
617 Scikit-learn: Machine learning in Python. *Journal of Machine Learning*
618 *Research*, 12:2825–2830, 2011.
- 619 [10] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Intro-*
620 *duction to Information Retrieval*. Cambridge University Press, New York,
621 NY, USA, 2008.
- 622 [11] Mahendra Sahare and Hitesh Gupta. A review of multi-class classifica-
623 tion for imbalanced data. *International Journal of Advanced Computer*
624 *Research*, 2(3), 2012.
- 625 [12] *NIST/SEMATECH e-Handbook of Statistical Methods*. NIST/SEMA-
626 TECH, April 2012.
- 627 [13] Andrea Volkens, Pola Lehmann, Theres Matthieß, Nicolas Merz, Sven Re-
628 gel, and Bernhard Weßels. The manifesto data collection. manifesto project
629 (mrg/cmp/marpor). version 2017b, 2017.

630 A Slides