

Neural network-algoritmes: een nieuw gereedschap voor de historicus?

Onderzoek met een neural network naar de houding ten opzichte van de middeleeuwen in Nederlandse krantenartikelen tussen 1830-1910

RICK SCHOUTEN
5717094

BA Scriptie Geschiedenis (15 ECTS)

Begeleider: dr. A.C van Rhijn

24 juli 2020
Universiteit Utrecht

Neural network-algoritmes: een nieuw gereedschap
voor de historicus?

Samenvatting

In deze scriptie heb ik onderzocht hoe tussen 1830-1910 de houding binnen de Nederlandse samenleving ten opzichte van de middeleeuwen ontwikkelde, waarbij ik gebruik heb gemaakt van gedigitaliseerde krantenartikelen van verschillende zuilen of politieke/religieuze stromingen uit deze periode, waarin het woord ‘middeleeuwen’ (in verschillende vormen) voorkomt. Deze artikelen heb ik geclassificeerd met een *neural network* (een CNN-BiLSTM model), op of ze een negatieve, neutrale of positieve houding ten opzichte van de middeleeuwen reflecteerden; dit heb ik later echter moeten terugbrengen tot twee categorieën, een negatieve en niet-negatieve categorie.

De keuze voor, en verkenning van deze methode (deel één) staat samen met het onderzoek naar de negentiende-eeuwse middeleeuwen (deel twee) centraal. Ik betoog in het eerste deel onder andere dat historici de toepassing van een geautomatiseerde analyse met *neural network*-algoritmes zouden moeten omarmen, om zo de kansen die de toenemende beschikbaarheid van gedigitaliseerde archieven biedt, te kunnen benutten.

In het tweede deel presenteer ik de resultaten van de analyse door het algoritme van bijna 30.000 artikelen, en blijken er voor 1880 drie pieken van negativiteit ten opzichte van de middeleeuwen te zijn, waarin de katholieken vaak een tegenovergestelde trend lieten zien. Daarbij lijkt de ontwikkeling van de houding van katholieken en liberalen beïnvloed te zijn door enkele belangrijke publicaties in de negentiende eeuw, en ook door elkaar. Verder onderzoek zou zich kunnen richten op de onderwerpen waarover meer of minder negatief geschreven wordt.

Inhoud

Inleiding	9
Deel Eén: Methode.....	13
1. Introductie	15
2. <i>Digital Humanities</i> -onderzoek in verleden, heden en nabije toekomst	17
2.1. Verleden.....	17
2.2. Heden.....	20
2.3. Nabije toekomst.....	21
3. Problemen en bezwaren	23
4. Het model.....	29
4.1. Neural networks.....	29
4.2. Het CNN-BiLSTM model	30
4.3. De trainingsdata.....	33
4.4. Prestaties.....	34
4. Conclusie.....	39
Deel Twee: Resultaten	41
5. Introductie	43
6. Ruwe data.....	45
7. De middeleeuwen per zuil	51
7.1. Data en analyse.....	51
7.2. Verklaringen	57
8. Conclusie.....	61
Conclusie.....	63
Bibliografie.....	67
Appendices	71
A. Codelijst	72
B. Model	75

Inleiding

MUNSTER, den 1 Julij. In de verlopen week heeft alhier een voorval plaats gehad, dat eenige overeenkomst heeft met de Godsoordeelen uit de middeleeuwen. Twee jonge officieren, de baron v. Denkhaus en de heer v. Bounhart, beide luitenants, hadden, terwijl zij in een kofijhuis billard speelden, een hevigen twist gehad, waarbij de eerste zich eenige beledigende woorden jegens zijn medespeler had laten ontvallen. Daar deze woorden en op eene publieke plaats waren uitgesproken, [...] meende de heer v. Bounhart gedwongen te zijn in het openbaar voldoening te eischen en vervolgde hij den heer v. Denkhaus, wegens beledigingen voor het eergeregtshof te Munster.

Dit geregtshof stelde overeenkomstig de wet alles in het werk om den belediger over te halen zijne beledigende uitdrukkingen in te trekken, en, hiertoe niet kunnende geraken, velde het een vonnis, waarbij bepaald werd, [...] dat er tusschen beide partijen een duel volgens militaire regelen zoude plaats hebben.¹

Op 14 juli 1848 kregen de lezers van de *Leeuwarder Courant*, het bovenstaande bericht onder ogen. Het beschrijft een, kennelijk ook voor de negentiende-eeuwse Nederlander, nogal absurde situatie, waarin de wellicht herkenbare situatie van een hand gelopen biljardpartij, op volstrekt onherkenbare wijze eindigt met een veroordeling tot een duel door nota bene een eergerechtshof.

De lezer wordt hier twee dingen duidelijk; enerzijds dat de Duitsers er nogal vreemde, en achterhaalde wetten en gebruiken op na houden, en anderzijds dat deze vreemde manier van

¹ 'Duitschland', *Leeuwarder courant* (Leeuwarden 14 juli 1846) 1 (Geraadpleegd op 3 juli 2020 via Delpher, <http://resolver.kb.nl/resolve?urn=ddd:010583417:mpeg21:p001>).

handelen ook in de middeleeuwen aan de orde van de dag was. Vooral door die laatste kwalificatie laat blijken dat de schrijver van het artikel en hoogst waarschijnlijk ook de lezer hier ver boven staan. Beiden leven immers in de rationele, en bovenal beschaafde moderne tijd, waarin voor een institutie als het middeleeuwse godsoordeel, of de willekeur van een middeleeuws duel als genoegdoening voor gekrenkte eer, geen ruimte is.

Dit artikel reflecteert daarmee een nogal negatief beeld van de middeleeuwen, iets wat bij herhaling in dezelfde krant waarschijnlijk onbewust ook overgedragen wordt op de lezers. Tenminste, als de lezers er zelf al niet zo over dachten. Lezers van een krant zullen zich vaak toch tenminste enigszins kunnen herkennen in het wereldbeeld of normen en waarden van een krant. Vooral bij kranten met een duidelijke identiteit, zoals katholieke of protestantse kranten zal dit het geval zijn.

In dit onderzoek zal ik daarvan gebruik maken, als ik dit soort krantenartikelen waarin de middeleeuwen worden genoemd of gebruikt analyseer, om te onderzoeken hoe in het Nederland van de negentiende eeuw, de houding tot de middeleeuwen zich ontwikkelde.

Met dit onderzoek sluit ik daarmee aan op eerder onderzoek door Peter Raedts naar ontwikkeling van de middeleeuwen in onder andere de negentiende eeuw, waarbij de focus overigens met name lag op katholieken en protestanten.²

Volgens Raedts hebben de middeleeuwen in Nederland altijd een wat ambigue positie gehad, omdat er in het nationale verhaal van het protestantse Nederland maar weinig ruimte was voor de katholieke middeleeuwen. Nederland werd immers een onafhankelijk land door zich juist af te zetten tegen deze eeuwen van onderdrukking, toen de Nederlanders vonden dat hun katholieke Spaanse koning te ver ging, met de hevige vervolging van protestantse Nederlanders. In tegenstelling tot landen als Engeland, Frankrijk en Duitsland, waren er in Nederlandse middeleeuwse geschiedenis bovendien maar weinig aanknopingspunten voor een glorieus verhaal over de eigen natie. Voor het Nederlandse succesverhaal kon men immers beter terecht in de zeventiende-eeuwse Gouden Eeuw.

Dit was althans volgens Peter Raedts, voor de meerderheid van de Nederlandse bevolking het beeld van het verleden. De Nederlandse katholieken ontwikkelden wél een positiever beeld over de middeleeuwen, die deels gedreven werd door het zoeken naar een eigen identiteit, en deels ook als reactie op protestantse schrijvers die de middeleeuwen gebruikten of misbruikten in hun pogingen een Nederlandse geschiedenis te schrijven.

² Peter Raedts, *De ontdekking van de Middeleeuwen* (Amsterdam 2011) 227-276.

Raedts onderzoek was voornamelijk gebaseerd op een kwalitatieve analyse van belangrijke publicaties uit de negentiende eeuw over de middeleeuwen en tijdschriften. In dit onderzoek zal ik gebruik maken van alle krantenartikelen met een vermelding van de middeleeuwen uit in Nederland verspreide kranten tussen 1830 en 1910, die op 23 april via Delpher, het online krantenarchief van de Koninklijke Bibliotheek, beschikbaar waren.³ Voor de analyse van deze bijna dertig duizend artikelen, zal ik gebruik maken van een *neural network*-algoritme die de artikelen classificeert op of ze positief, negatief of neutraal over de middeleeuwen spreken. Zoals kranten zijn te onderscheiden naar bijvoorbeeld zuil of verspreidingsgebied, en uiteraard heel precies naar een moment in de tijd, zo is dat daarmee ook voor data die het classificatiemodel oplevert. Dit is waarom juist kranten zo'n interessante bron zijn om deze ontwikkeling te onderzoeken.

Naast dat ik in deze scriptie een poging zal doen het door Raedts geschetste beeld te testen en uit te breiden naar andere zuilen dan alleen de katholieke en protestantse, is ook de gekozen methode een groot en belangrijk onderdeel van dit onderzoek. Het onderzoek naar de negentiende-eeuwse middeleeuwen is in zekere zin een test casus voor de toepassing van *neural networks* in de analyse van historische krantenartikelen om de houding tot een bepaald onderwerp te onderzoeken.

De vraag die dan ook centraal staat in dit onderzoek is: hoe ontwikkelde tussen 1830-1910 de houding binnen de Nederlandse samenleving ten opzichte van de middeleeuwen, afgeleid uit het gebruik van de middeleeuwen in krantenartikelen uit deze periode, en hoe is deze ontwikkeling te verklaren, en hoe zijn *neural networks* door historici in te zetten in een dergelijk onderzoek?

Deze vraag zal ik in twee delen beantwoorden. In het eerste deel van dit onderzoek staat de methode centraal, en stel ik de vraag hoe *neural network*-algoritmes zijn in te zetten door historici bij het doen van onderzoek naar de ontwikkeling in de houding van een samenleving ten opzichte van een bepaald onderwerp, in dit geval de middeleeuwen, en hoe dit is te rechtvaardigen. In het tweede deel van deze scriptie richt ik mij vervolgens op de resultaten van het classificatie-model, en geef ik antwoord op de vraag hoe de houding ten opzichte van de middeleeuwen zich ontwikkelde in de verschillende zuilen in de periode 1830-1910.

³ Delpher, <https://www.delpher.nl/nl/kranten> , alle artikelen gevonden met de zoekterm 'middeleeuw*'

Deel Eén: Methode

1. Introductie

Voor dit onderzoek heb ik zoals ik in de inleiding schreef, gekozen om alle gedigitaliseerde krantenartikelen op Delpher uit de onderzochte periode als bron te gebruiken. Daarmee rijst de vraag hoe die enorme hoeveelheid krantenartikelen geanalyseerd moeten en kunnen worden, of dat er misschien maar een beperkt aantal artikelen in het onderzoek meegenomen kan worden. Zoals uit de onderzoeksvraag al blijkt heb ik gekozen voor een algoritme om de artikelen te analyseren, maar die keuze dient uiteraard wel onderbouwd en verdedigd te worden met goede argumenten, waarbij ook alternatieve methoden worden afgewogen.

In dit eerste deel zal ik daarom voornamelijk stilstaan bij de gekozen methode. Daarbij zal ik in het eerste hoofdstuk kort de historische en technische ontwikkeling van het geautomatiseerd analyseren van historische bronnen bespreken. Hierna zal ik in het tweede hoofdstuk ingaan op door historici opgeworpen problemen en bezwaren tegen het gebruik van een gedigitaliseerde en geautomatiseerde analyse van bronnen. In het laatste en derde hoofdstuk zal ik vervolgens uitleggen hoe het model voor dit onderzoek werkt en hoe ik het heb getraind, en wat de tekortkomingen van het model zijn. Aan het slot van dit eerste deel volgt een conclusie waarin ik antwoord geef op de deelvraag die in dit hoofdstuk centraal staat, namelijk hoe *neural network*-algoritmes zijn in te zetten door historici bij het doen van onderzoek naar de ontwikkeling in de houding van een samenleving ten opzichte van een bepaald onderwerp, in dit geval de middeleeuwen, en hoe dit te rechtvaardigen is.

2. *Digital Humanities*-onderzoek in verleden, heden en nabije toekomst

In dit hoofdstuk zal ik zoals aangekondigd ingaan op de door mij gekozen methode. Allereerst zal ik een kort historisch overzicht geven van de *digital humanities* waaronder iedere vorm van gedigitaliseerd geesteswetenschappelijk onderzoek valt, en de toepassing van digitale (geautomatiseerde) methoden op de analyse van tekstuele bronnen. Hieruit zal blijken dat de door mij gekozen methode past in deze langere ontwikkeling, en op deze manier toegepast, mogelijk een nieuwe volgende stap hierin kan zijn. Vervolgens zal ik de door mij gekozen methode vergelijken met alternatieve methodes om een grote hoeveelheden bronnen, in dit geval krantenartikelen, te analyseren.

2.1. Verleden

Het gebruik van computers in historisch of geesteswetenschappelijk onderzoek, is niet iets nieuws. Sinds het bestaan van computers vanaf de Tweede Wereldoorlog, zijn er al geesteswetenschappers bezig met het gebruiken van computers als hulpmiddel in hun onderzoek. Het was alleen niet een historicus, maar een theoloog en Jezuiet, Roberto Busa (1913-2011), die hiermee de eerst was.⁴ Ter verkrijging van een doctoraat aan de pauselijke

⁴ Shawn Graham, Ian Milligan en Scott Weingart, *Exploring big historical data : the historian's macroscope* (Londen 2016) 20–21.

universiteit in Rome, had hij een concordantie gemaakt van de werken van Thomas van Aquino. Dit is een lijst waarbij van ieder woord, voor iedere keer dat het voorkomt in het corpus, in de context wordt weergegeven. Het idee is dat het hierdoor gemakkelijk is te zien hoe een bepaald woord in een bepaald werk, of in bepaalde werken wordt gebruikt, zonder iedere keer opnieuw alle werken opnieuw door te hoeven spitten. Dit monnikenwerk had Busa volledig handmatig uitgevoerd, maar in de computer zag hij een hulpmiddel die dit sneller zou kunnen. Zodoende melde hij zich bij IBM met de vraag of zij zoiets voor hem konden ontwikkelen. Hoewel medewerkers van het bedrijf dit voor onmogelijk hielden, zag Thomas Watson, de directeur, wel de potentie van dit idee, waarna het project in de hierop volgende dertig jaar succesvol werd afgerond.⁵

Na Busa zijn er meer geesteswetenschappers geïnteresseerd geraakt in de toepassing van computers in hun onderzoek, ook historici. Rond de jaren '70 waren er al enkele tientallen historische artikelen en onderzoeken gepubliceerd waarin gebruik werd gemaakt van zulke onderzoeksmethodes.⁶ De door de *Annales*-school gestimuleerde interesse in geschiedenis op macroniveau, gaf door de verwerking van grote hoeveelheden kwantitatieve data die daarvoor nodig was, een impuls aan digitaal historisch onderzoek.⁷ Met de komst van de nieuwe technologie, kwamen dus ook nieuwe bronnen in trek, zoals bevolkings- en economische statistieken. Het met computers onderzoeken van teksten bleef ook na de komst van de PC, echter voornamelijk besteed aan literatuur- en taalwetenschappers, die voornamelijk allerlei tekstuele patronen en statistieken verzamelden.⁸ Door een gebrek aan goede programma's, technieken en überhaupt de staat van de technologie, was het voor historici tot voor kort lastig om tekstuele bronnen digitaal te onderzoeken; behalve dan het op een scherm bestuderen ervan en het makkelijk kunnen terugvinden van bepaalde zinnen of woorden.

Met de tijd zijn computers steeds beter en sneller geworden waardoor het tegenwoordig mogelijk is om Busa's concordantie binnen enkele minuten te creëren, inclusief het bij elkaar verzamelen en downloaden van alle teksten. Want ook steeds meer bronnen en archieven zijn

⁵ Ibidem.

⁶ Robert P. Swierenga, 'Clio and computers: A survey of computerized research in history', *Computers and the Humanities* 5 (1970) 20-21. Robert P. Swierenga, 'Clio and computers: A survey of computerized research in history', *Computers and the Humanities* 5 (1970) 1-21, aldaar 1-21 <doi:10.1007/BF02404252>.

⁷ Graham, Milligan en Weingart, *Exploring big historical data : the historian's microscope*, 22-23.

⁸ Susan Hockey, 'The History of Humanities Computing', in: Susan Schreibman, Ray Siemens en John Unsworth ed., *A Companion to Digital Humanities* (Oxford 2004) 1-19.

inmiddels geheel of (vooralsnog) gedeeltelijk gedigitaliseerd, en online toegankelijk gemaakt. Hierdoor is het mogelijk om zonder al te veel moeite archieven te doorzoeken aan de andere kant van de wereld.

Deze ontwikkeling is voor de geesteswetenschappen, en met name de geschiedwetenschap van groot belang. Het verleden is immers alleen (ten dele) kenbaar door het gebruik van zoveel mogelijk primaire bronnen, zoals Leopold von Ranke leert.⁹ Hoe meer bronnen er toegankelijk zijn van alle verschillende kanten van een historische gebeurtenis of ontwikkeling, hoe beter en in zekere zin objectiever de geschiedschrijving daarover kan worden. Uiteraard is nog maar een deel van alle archieven gedigitaliseerd, maar dit probleem zal met de tijd kleiner worden. Het feit dat er nu al zoveel beschikbaar is, en zoveel meer te onderzoeken is dan voorheen, is al een enorme winst.

Anderzijds wordt hierdoor ook weer een nieuw probleem gecreëerd, want hoe kan van tevoren een goede selectie worden gemaakt uit de veelheid van deze bronnen? Natuurlijk is een historicus wel gewend om een selectie te maken, maar voordat er theoretische keuzes gemaakt worden, is een groot deel van de bronnen al afgevallen vanwege praktische en financiële redenen; ongepubliceerde bronnen die in een archief ver weg liggen, vallen bijvoorbeeld vrij snel af. Nu er veel meer bronnen toegankelijker zijn, wordt het bovendien lastiger om de juiste theoretische beslissingen te nemen, met name omdat onmogelijk alle bronnen van tevoren goed geëvalueerd kunnen worden. Een goed voorbeeld hiervan is het soort onderzoek wat ik hier doe, waarin de krant als bron of onderzoeksobject wordt gebruikt¹⁰.

Daarbij maakt deze overvloed aan toegankelijke bronnen het gedeelte van de bronnen wat gebruikt kan worden voor traditioneel historisch onderzoek relatief veel kleiner ten opzichte van de hoeveelheid toegankelijke bronnen. Daarom, maar vooral om van het potentieel gebruik te kunnen maken wat deze grote hoeveelheden gedigitaliseerde bronnen bieden, is een nieuwe methode voor het analyseren van bronnen noodzakelijk.

⁹ Andreas Boldt, 'Perception, Depiction and Description of European History: Leopold von Ranke and his Development and Understanding of Modern Historical Writing', *eSharp* (2007) 1–17, aldaar 13–14.

¹⁰ Joris van Eijnatten, Toine Pieters en Jaap Verheul, 'Big Data for Global History. The Transformative Promise of Digital Humanities', *BMGN - Low Countries Historical Review* 128 (2013) 55–77, aldaar 59.

2.2. Heden

Omdat ik onmogelijk kan ingaan op alle soorten bronnen en alle manieren van analyseren die daarbij mogelijk zijn, en bovendien dat niet het doel van dit onderzoek is, zal ik mij hier alleen richten op methodes voor het analyseren van historische kranten.

De gangbare methode voor krantenonderzoek is de zogenaamde *content analysis*, of inhoudsanalyse.¹¹ Hiermee wordt kwantitatieve data verzameld over de inhoud van in dit geval kranten, waarmee een ontwikkeling gedurende een bepaalde tijd inzichtelijk kan worden gemaakt.¹² De onderzoeker stelt van tevoren een codelijst op met variabelen die hij of zij wil onderzoeken, zoals journalistieke genres, waarna door een selectie van kranten wordt gebladerd en vervolgens de voor het onderzoek relevante artikelen worden gecodeerd.¹³ Dit arbeidsintensieve werk wordt soms deels uitbesteed aan andere codeurs, die hiervoor een training van zo'n dertig uur ondergaan om een acceptabele mate van *intercoder reliability* te bereiken, dat wil zeggen, overeenstemming met de onderzoeker en met elkaar.¹⁴

Naast de selectie van krantentitels, wordt vanuit praktische overweging ook een selectie gemaakt in de edities van de kranten die onderzocht gaan worden. Doorgaans wordt deze tweede selectie gedaan door een vaste dag in een vaste week in een vaste maand te kiezen, zoals iedere dinsdag in de derde week van maart, of iedere woensdag voor iedere tweede week van iedere maand.¹⁵ De selectie die hieruit vervolgens ontstaat heet een *constructed week sample*.¹⁶ Op deze manier wist Rutger de Graaf voor zijn proefschrift in drie jaar, samen met zes andere codeurs, ruim dertigduizend artikelen te coderen.¹⁷

Naast het handmatig coderen zijn er ook technieken om dit, of ander soortgelijk onderzoek, (gedeeltelijk) geautomatiseerd te doen. De meest eenvoudige manier is om te tellen hoe vaak bepaalde woorden, of woordcombinaties binnen een bepaalde afstand, voorkomen in één of meerdere bronnen.¹⁸

¹¹ R.H. de Graaf, 'Journalistiek in beweging. Veranderende berichtgeving in kranten en pamfletten (Groningen en 's-Hertogenbosch 1813-1899)' (Proefschrift Utrecht 2009) 11-13.

¹² Ibidem.

¹³ Kimberly A. Neuendorf, *The content analysis guidebook* (2de druk; Los Angeles 2017) 40.

¹⁴ Idem, 221,224.

¹⁵ De Graaf, 'Journalistiek in beweging', 346-347.

¹⁶ Ibidem.

¹⁷ Idem, 13,348.

¹⁸ Van Eijnatten, 'Big Data for Global History', 62-65.

Meer geavanceerde manier is door gebruik te maken van *machine learning*-algoritmes. Dit soort algoritmes zijn er in vele soorten en maten, maar ze zijn op basis van de manier waarop ze worden getraind, grofweg in te delen in twee verschillende types: *supervised*- en *unsupervised*- algoritmes.¹⁹ Dit laatste type algoritmes classificeert de data (dit kunnen ook teksten zijn) zelfstandig op basis van overeenkomsten die het in de data opmerkt, in een van te voren gekozen aantal categorieën. Het grote nadeel daarvan is echter dat het niet mogelijk is om vooraf inhoudelijk te bepalen in welke categorieën de data moet worden ingedeeld. Overigens is dat ook juist de kracht van dit algoritme, omdat het verrassende inzichten kan bieden en onbekende patronen in een corpus zichtbaar kan maken.²⁰

Bij *supervised*-algoritmes zijn de categorieën van tevoren wél duidelijk. Het algoritme leert op basis van artikelen waarvan de categorieën bekend zijn. Het nadeel van veel van deze algoritmes is dat ze, net als de *unsupervised*-algoritmes, de tekst zien als een *bag-of-words*.²¹ Dit betekent dat de tekst niet als zinnen (of als sequentie van woorden) het model in gaat, maar als een soort frequentietabel, waarbij voor ieder woord staat aangegeven hoe vaak het in de tekst voorkomt. Woorden worden hierbij vaak teruggebracht tot de stam, en in sommige gevallen, zelfs tot de woordenboekvorm (lemmatisering) of de tegenwoordige tijd²². Veel semantische informatie gaat hierbij zo dus verloren.²³ Waar dit in sommige toepassingen voor het classificeren van langere documenten niet noodzakelijk een probleem is, is dat het wel voor het classificeren van kortere teksten waarbij het vaak aankomt op subtiele verschillen of nuances zoals in dit onderzoek.²⁴

2.3. Nabije toekomst

Voor dit onderzoek maak ik daarom gebruik van *supervised*-algoritmes die teksten wél als een sequentie van woorden zien. Het type algoritmes die dit kunnen zijn *neural networks*, hoewel niet ieder type *neural network* dit kan, maar daar ga ik later op in. Het voordeel van dit soort algoritmes dat ze door de manier waarop ze teksten 'lezen', veel beter dan andere

¹⁹ Matthew L. Jockers en Ted Underwood, 'Text-Mining the Humanities', in: Susuan Schreibman, Ray Siemens en John Unsworth ed., *A new Companion to Digital Humanities* (Oxford 2016) 291–306, aldaar 293.

²⁰ Ibidem.

²¹ Ibidem.

²² Jockers en Underwood, 'Text-Mining the Humanities', 299.

²³ Marcel Broersma en Frank Harbers, 'Exploring Machine Learning to Study the Long-Term Transformation of News: Digital newspaper archives, journalism history, and algorithmic transparency', *Digital Journalism* 6 (2018) 1150–1164, aldaar 1156 <doi:10.1080/21670811.2018.1513337>.

²⁴ Ibidem.

algoritmes, en vaak bijna net zo goed als mensen, complexere classificatieproblemen kunnen oplossen.

Los van dit specifieke type algoritme, heeft van een analysemethode als deze een aantal belangrijke voordelen. Allereerst kan er een gigantische hoeveelheid bronnenmateriaal over een veel langere periode geanalyseerd en inzichtelijk gemaakt kan worden, en dit alles kan bovendien in een relatief beperkte tijd, door zelfs één persoon. Het is daarnaast ook gemakkelijk om dit soort onderzoek uit te breiden, of te 'updaten', als er weer nieuw bronnenmateriaal beschikbaar komt; deze bronnen hoeven slechts door het al bestaande model heen gehaald te worden. Hierdoor kunnen veel grotere vragen, over veel langere periodes en/of voor grotere geografische gebieden beantwoord worden.

Misschien wel de belangrijkste toevoeging van deze methode aan de geschiedwetenschap, is dat met deze methode onzichtbare en onbekende ontwikkelingen in tijd en ruimte, en vooral ook uitzonderingen daarop, aan het licht gebracht kunnen worden die anders mogelijk voor lange tijd verborgen zouden zijn gebleven. Deze opmerkelijkheden kunnen vervolgens dienen als aanknopingspunt voor dieper kwalitatief onderzoek, waardoor er uiteindelijk een completer beeld van het verleden worden gereconstrueerd.

Naast deze voordelen, zitten er aan deze manier van werken ook nadelen of beperkingen. Die zijn dan ook door critici tegen het gebruik van een dergelijke methode ingebracht. Een van de meest genoemde bezwaren en problemen die betrekking hebben op het type algoritme dat ik hier gebruik, zal ik daarom in het volgende hoofdstuk bespreken en waar nodig weerleggen.

3. Problemen en bezwaren

In de literatuur zijn grofweg drie typen bezwaren te onderscheiden. Het eerste is vooral van praktische aard, namelijk dat deze manier van onderzoek doen technische vaardigheden vereist waarover historici niet beschikken. Het tweede type is van meer theoretische aard, en gaat in op de claim van volledigheid, tenminste de schijn daarvan die kan worden gewekt met deze nieuwe manier van onderzoek doen. Ten slotte zijn er nog technische bezwaren, hoewel die deels ook van theoretische aard zijn, en betrekking hebben op het gebrek aan inzichtelijkheid van *neural networks*, en technische of praktische problemen vanwege fouten in de in digitale tekst omgezette bronnen.

Allereerst dus de praktische problemen of bezwaren tegen het gebruik van algoritmes in historisch onderzoek. Eén van de grootste problemen die vaak genoemd wordt, is dat aangezien de historicus geen informaticus is, hij of zij afhankelijk is van experts of van (inflexibele) *tools* die door deze experts zijn gemaakt, op het gebied van bijvoorbeeld *machine learning*.²⁵ Hoewel dat an sich geen probleem is, wijzen critici op de verschillende belangen, prioriteiten en een wederzijds gebrek aan kennis over het vakgebied van de ander, met een haperende samenwerking tot gevolg.²⁶ Bovendien wordt een onderzoek zo al snel een

²⁵ Broersma en Harbers, 'Exploring Machine Learning', 1157.

²⁶ Huub Wijfjes, 'Digital Humanities and Media History. A Challenge for Historical Newspaper Research',

groot interdisciplinair project, terwijl de historicus juist gewend is om vaak alleen, of met slechts een paar vakgenoten samen te werken.²⁷

Een oplossing voor zowel de afhankelijkheid van, als de stoeve samenwerking met experts, is in zekere zin eenvoudig: de historicus kan zich immers ook zélf de benodigde kennis en vaardigheden aanleren. Waar het programmeren en trainen van algoritmes tot enkele jaren geleden toebehoorde tot de wereld van de computernerds, met complexe programmeertalen, ingewikkelde terminologie en wiskundige theorieën, is dat tegenwoordig steeds minder het geval. De wiskundige vergelijkingen die schuil gaan achter de algoritmes, of in feite de algoritmes vormen, zullen misschien altijd ondoorgrondelijk blijven, maar dat is geen voorwaarde voor het gebruik van deze technieken. De essentie van wat een algoritme doet is in principe voldoende om het te kunnen toepassen; en dat is juist veel gemakkelijker te begrijpen.

Ook het leren van de benodigde programmeervaardigheden lijkt een grotere hindernis dan deze daadwerkelijk is. De grootste programmeertaal voor *machine learning*, Python, is vrijwel de gemakkelijkste taal om te leren. Het internet staat bovendien vol met (gratis) handleidingen, cursussen en video's, waarin alles tot in detail wordt uitgelegd, en zijn er tal van boeken die gericht zijn op beginners.²⁸ Met andere woorden, wie geen gebruik wil maken van experts om eerder genoemde redenen, maar wel de toegevoegde waarde en de mogelijkheden van *machine learning* ziet, hoeft slechts tijd te steken in het leren ervan. En mocht de historicus tijdens zijn onderzoek dan toch nog de hulp moeten inschakelen van een expert, dan zal die samenwerking en communicatie daardoor veel soepeler verlopen.

Het tweede bezwaar, is een theoretisch punt, gebaseerd op verkeerde verwachtingen, die ofwel zijn gecreëerd door de digitale historici, ofwel door tegenstanders daarop zijn geprojecteerd. Zo is een argument voor het gebruik van een geautomatiseerde analyse dat dit een volledig en objectief beeld van het verleden kan geven dan traditioneel onderzoek.²⁹ Daar wordt terecht tegenin gebracht dat natuurlijk ook deze vorm van geschiedschrijving beperkt

Tijdschrift voor Mediageschiedenis 20 (2017) 4–24, aldaar 8,17. En: Eijnatten, Pieters, en Verheul, 'Big Data for Global History', 60.

²⁷ Eijnatten, Pieters, en Verheul, 'Big Data for Global History', 60.

²⁸ Bijvoorbeeld: Aurélien Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems* (Boston 2019).

²⁹ Onder andere: Graham, Milligan en Weingart, *Exploring big historical data*, 32-33. En: Broersma en Harbers, 'Exploring Machine Learning' 1152, 1157.

zal blijven. De historicus blijft immers afhankelijk van gekleurde bronnen die zijn overgeleverd, en de subjectiviteit die in die overlevering zelf zit.³⁰ Maar dit is daarmee geen reden om deze manier van onderzoek ter geheel zijde te schuiven, aangezien deze problemen voor kleinschalig kwalitatief onderzoek net zo geldig zijn. Hierin wordt zelfs van die overgeleverde gekleurde bronnen een kleinere selectie wordt gemaakt, die voor een groot deel ook niet eens zo zeer op theoretische overwegingen zijn gebaseerd. Dit argument pleit daarmee dus ofwel alleen maar vóór het gebruik van digitale analyses (tenminste als aanvulling), ofwel voor het helemaal stoppen met geschiedschrijving.

Maar ook bij digitaal onderzoek wordt er in veel gevallen gebruik gemaakt van een selectie van al het overgeleverde bronnenmateriaal, namelijk dat deel wat gedigitaliseerd en toegankelijk is.³¹ Zolang er echter zoals bij ieder historisch onderzoek rekening wordt gehouden met die beperktheid, en een eventuele selectie van bronnen voor digitalisering door het archief zorgvuldig is gedaan, is dit geen probleem.³² Los daarvan is de hoeveelheid digitaal bruikbare bronnen, ook al is dit een selectie, vaak alsnog heel veel meer dan een historicus op analoge wijze ooit zou kunnen gebruiken. Dit onderzoek was bovendien zonder gedigitaliseerde kranten praktisch onmogelijk geweest, ongeacht de methode van analyse, omdat het ondoenlijk is om op analoge wijze, zelfs uit een beperkte selectie kranten, alle artikelen met het woord 'middeleeuwen' erin te verzamelen.

Onder het derde en laatste type, technisch-theoretische problemen vallen drie veel genoemde problemen of aandachtspunten. Het eerste heeft te maken met de veranderlijkheid van taal door de tijd heen, waarmee een model die getraind is over een langere periode geen rekening houdt. Dit probleem kan enerzijds eenvoudig worden opgelost door voor verschillende periodes verschillende modellen te trainen.³³ Anderzijds is het echter nog maar de vraag in hoeverre dit in de praktijk nu echt een probleem is. In bijvoorbeeld dit onderzoek waarin periode van tachtig jaar vanaf 1830 wordt onderzocht, zullen er weinig relevante woorden zijn die ten opzichte van het begin van de periode iets heel anders betekenden, dan aan het einde van die periode. Bovendien zal als de betekenis van een woord veranderd, de context waarin die dat woord gebruikt wordt dat ook doen. Omdat een model zijn voorspelling niet op

³⁰ Eijnatten, Pieters, en Verheul, 'Big Data for Global History', 60.

³¹ Wijfjes, 'Digital Humanities and Media History', 17.

³² Jockers en Underwood, 'Text-Mining the Humanities', 300-301.

³³ Broersma en Harbers, 'Exploring Machine Learning' 1156.

één enkel woord baseert maar op de woorden in hun context, zal het daadwerkelijke probleem in de praktijk zeer beperkt zijn. Mocht het toch een probleem zijn, dan ligt waarschijnlijk eerder aan de trainingsdata die te eenzijdig is, dan aan de technologie zelf.

Het tweede probleem is de door het *Optical Character Recognition*-algoritme (OCR) verkeerd 'gelezen' artikelen. Zoals bij alle historische bronnen, gedigitaliseerd of niet, zijn ook de gescande kranten soms moeilijk leesbaar zijn vanwege verwerking door de tijd, moeilijk leesbare lettertypes, vouwen of inktvlekjes. De OCR probeert van dat wat het vermoed een letter te zijn, toch iets te maken, ook al is het maar half zichtbaar of een inktvlek. Hierdoor kunnen verkeerde of losse tekens in een tekst, of binnen woorden zelf verschijnen.³⁴

Hoewel dit problematisch lijkt voor een geautomatiseerde beoordeling van artikelen, blijkt dit in de praktijk erg mee te vallen. Hoewel de accuraatheid van een algoritme inderdaad iets omlaag gaat, blijkt in de praktijk het probleem veel kleiner dan wordt voorgesteld door sceptici.³⁵ Vooral bij gebruik van een *neural network* blijkt de schade aan de accuraatheid van het model zeer beperkt, aangezien, net als bij de veranderende woorden, de context waarin woorden staan ook wordt meegenomen.³⁶ Als het model bovendien ook wordt getraind op dit soort artikelen, dan helpen die zelfs om überhaupt een sterker model te maken.³⁷ Daarbij is de OCR-technologie constant in ontwikkeling, evenals de correctiealgoritmes die steeds beter in staat zijn om OCR-fouten te herstellen.³⁸ Dit probleem zal in de toekomst dus in de nabije toekomst steeds kleiner worden.

Het derde en laatste belangrijke technisch-theoretisch probleem wat ik hier bespreek, is het bezwaar dat algoritmes, en *neural networks* in het bijzonder, een zogenaamde *black box* zijn: het is onbekend op basis waarvan precies een specifiek krantenartikel door een model wordt

³⁴ Jockers en Underwood, 'Text-Mining the Humanities', 300.

³⁵ Emilia Apostolova en R Andrew Kreek, 'Training and Prediction Data Discrepancies: Challenges of Text Classification with Noisy, Historical Data', in: *Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text* (Brussel 2018) 104–109, aldaar 104–109 <doi:10.18653/v1/w18-6114>.

³⁶ Ibidem. En: Carolyn Strange e.a., 'Mining for the Meanings of a Murder: The Impact of OCR Quality on the Use of Digitized Historical Newspapers', *Digital Humanities Quarterly* 8 (2014).

³⁷ Apostolova, 'Training and Prediction', 104–109.

³⁸ Bijvoorbeeld: Janneke van der Zwaan en Lotte Wilms, 'Ochre. A Toolbox for OCR Post-Correction' (Koninklijke Bibliotheek, 2018) <doi:http://doi.org/10.5281/zenodo.1189245>. En: Kimmo Kettunen, 'Ground Truth OCR Sample Data of Finnish Historical Newspapers and Journals in Data Improvement Validation of a re-OCRing Process', *Liber Quarterly* 30 1–20 <doi:10.18352/lq.10322>.

geclassificeerd in een bepaalde categorie.³⁹ Hierdoor kan het in theorie zo zijn dat een model op basis van schijnbaar irrelevante woorden of woordcombinaties een tekst tijdens het trainingsproces per ongeluk indeelt in de juiste categorie. Hoewel dit op het eerste gezicht een valide bezwaar lijkt, is hiertegen toch een en ander in te brengen.

Ten eerste is het gebrek aan inzichtelijkheid niet noodzakelijk problematisch, omdat niet alleen een algoritme black box is, maar dat wat het vervangt ook: het menselijk brein. Zou iemand alle artikelen zelfstandig hebben gelezen en geclassificeerd, dan hoeft diegene niet te verantwoorden op basis van precies welke woorden hij of zij heeft besloten een specifiek artikel in een bepaalde categorie te plaatsen. Het is bovendien ook maar de vraag dat überhaupt mogelijk is voor ons mensen, aangezien wij vaak onbewust in een fractie van een seconde iets kunnen beoordelen als bijvoorbeeld positief of negatief. Als iemand die beslissing dan achteraf gaat verklaren, gaat iemand in feite proberen vanuit die beslissing te gissen naar wat het was dat het onbewuste tot die beslissing heeft doen komen. Of dat dan ook echt de reden was voor die beslissing blijft voor altijd onbekend, zelfs voor het individu, ook al lijkt er een logische verklaring te zijn.

Ten tweede is het belangrijkste dat het model voor een substantieel deel de artikelen juist classificeert. Hoe het precies tot die beoordelingen komt is in feite niet per se van belang om te weten, ook al zijn er bepaalde zwakheden in dat model. Mensen maken ook fouten, al zijn die vaak heel anders dan die van een model, waardoor de fouten van een model problematisch lijken. Daarbij is het voor een model na het trainen ervan duidelijk hoe goed het is, terwijl dit bij een individuele onderzoeker vaak onbekend blijft. In die zin is een model misschien juist wel transparanter dan een mens.

Ten slotte moet ik ook bij hierbij nog opmerken dat het black box-probleem iets is wat over niet al te lange tijd tot het verleden zal behoren, of althans sterk gereduceerd zal zijn. Dit probleem is een onderwerp waarnaar momenteel van onderzoek gedaan wordt, wat inmiddels ook al enkele oplossingen heeft opgeleverd.⁴⁰

³⁹ Broersma en Harbers, 'Exploring Machine Learning' 1157-1158.

⁴⁰ Bijvoorbeeld: Mengnan Du e.a., 'On attribution of recurrent neural network predictions via additive decomposition', *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019* (2019) 383–393. En: Yujia Zhang, e.a., "'Why Should You Trust My Explanation?'" *Understanding Uncertainty in LIME Explanations*' (2019). En: Yao Ming, Huamin Qu, en Enrico Bertini, 'RuleMatrix: Visualizing and Understanding Classifiers with Rules', *IEEE Transactions on Visualization and Computer Graphics* 25 (2019) 342–352.

Nu de belangrijkste problemen en bezwaren zijn besproken, en waar nodig toegelicht, afgezwakt of verworpen, blijken er geen grote onoverkomelijke bezwaren meer te zijn tegen het gebruik van een *neural network*-algoritme. In het volgende hoofdstuk zal ik mij daarom gaan richten op het model wat ik voor dit onderzoek heb gebruikt.

4. Het model

In deze paragraaf zal ik allereerst uitleggen hoe *neural networks* werken, hoe het algoritme voor dit onderzoek is opgebouwd, en hoe de trainingsdata tot stand is gekomen. Daarna zal ik ook ingaan op de zwakheden en beperkingen van het model, die gedeeltelijk voortkomen uit de categorieën waarvoor ik heb gekozen die uiteindelijk niet optimaal bleken, waarna ik tot slot ook de prestatiestatistieken van het model laat zien.

4.1. *Neural networks*

Een *neural network* bestaat uit verschillende lagen met daarin zogenaamde neuronen. De data stroomt vervolgens door die lagen, door die neuronen, door het model heen, waarbij iedere neuron een berekening maakt met de waarden die er instromen. Wat er in de neurons gebeurt wordt bepaald door de gekozen formule of *activation function* voor de neuronen in een bepaalde laag. De precieze parameters in die formule worden getraind tijdens het trainingsproces van het model. Niet iedere neuron krijg overigens alle data te zien, het kan ook klein een deel van de data zijn, bijvoorbeeld één waarde, maar dit wordt door het model zelf bepaald tijdens het trainen. De laatste laag van het model geeft het resultaat, of de voorspelling, voor hoe waarschijnlijk het bijvoorbeeld is dat de data in een bepaalde categorie past.

Voor het trainen van een model is trainingsdata nodig. Deze data bestaat uit gelabelde voorbeelden, waarbij één voorbeeld in dit geval één artikel is, en het label de categorie is waartoe het artikel behoort. Tijdens het trainen gaat er een kleinere set of *batch* van voorbeelden door het model heen, waarna het verschil wordt berekend, de *loss*, tussen het echte label, en het voorspelde of berekende label volgens het model. Op basis daarvan past het netwerk vervolgens van achter naar voor met een optimalisatiealgoritme de parameters van de neuronen iets aan, waarna dit proces zich herhaalt met volgende *batch*.

Een fundamenteel verschil tussen verschillende *neural networks*, zijn de type lagen met de neuronen die gebruikt worden. Hoewel de standaard neuronen zoals ik net beschreef vaak goed werken, zijn ze problematisch bij sequentiële voorbeelden, zoals teksten en chronologische data omdat het geen geheugen heeft. Als een tekst woord voor woord aan het model wordt gegeven, dan ziet het model ieder woord als een apart voorbeeld, waarvoor dan ook een aparte voorspelling gedaan. Maar, als de tekst wél in zijn geheel het model ingaat, dan wordt tekst in feite een *bag-of-words*, omdat het model geen rekening houdt met de volgorde van, en eventuele relatie tussen woorden.

Een *Recurrent Neural Network* (RNN) lost dit probleem op, aangezien het wél een geheugen heeft. De tekst stroomt woord voor woord door het netwerk, en iedere neuron gebruikt de uitkomst van de vorige berekening, van het vorige woord dus, in de nieuwe berekening, van het nieuwe woord. Deze opgeslagen uitkomst heet de *hidden state*. Dit is alleen wel een korte termijn geheugen, omdat het ieder nieuw woord net zo zwaar meetelt als alle voorgaande woorden samen.

4.2. Het CNN-BiLSTM model

In het model voor dit onderzoek heb ik gebruik gemaakt van een complexere RNN's met een langer geheugen: de *Long Short-Term Memory*-cel (LSTM). Naast een *hidden state*, wat een meer korte termijn geheugen is, heeft deze cel ook een zogenaamde *cell state*. Dit is een veel langere termijn geheugen met relatief belangrijke informatie uit alle voorgaande woorden uit een tekst. Of en hoeveel de informatie in deze *cell state* wordt aangepast door een nieuw woord, wordt bepaald door drie poorten (*gates*), die alle drie weer een klein *neural network* op zich zijn. Deze poorten reguleren wat er onthouden en vergeten wordt. Een cel leert uiteindelijk in de training wat belangrijke woorden en combinaties van woorden zijn, voor de

beoordeling van een tekst. Deze worden als het ware onthouden tot het einde van de tekst, waarna op basis daarvan een voorspelling wordt gedaan.

Een variant op de LSTM-laag is een *Bi-directional LSTM* (Bi-LSTM). Hierbij gaat de tekst niet alleen van voor naar achter, maar ook van achter naar voor door de cellen. Hierdoor kan het model nog beter rekening houden met woorden en frases die later in de tekst voorkomen, maar die de betekenis van woorden veel eerder in de tekst veranderen.

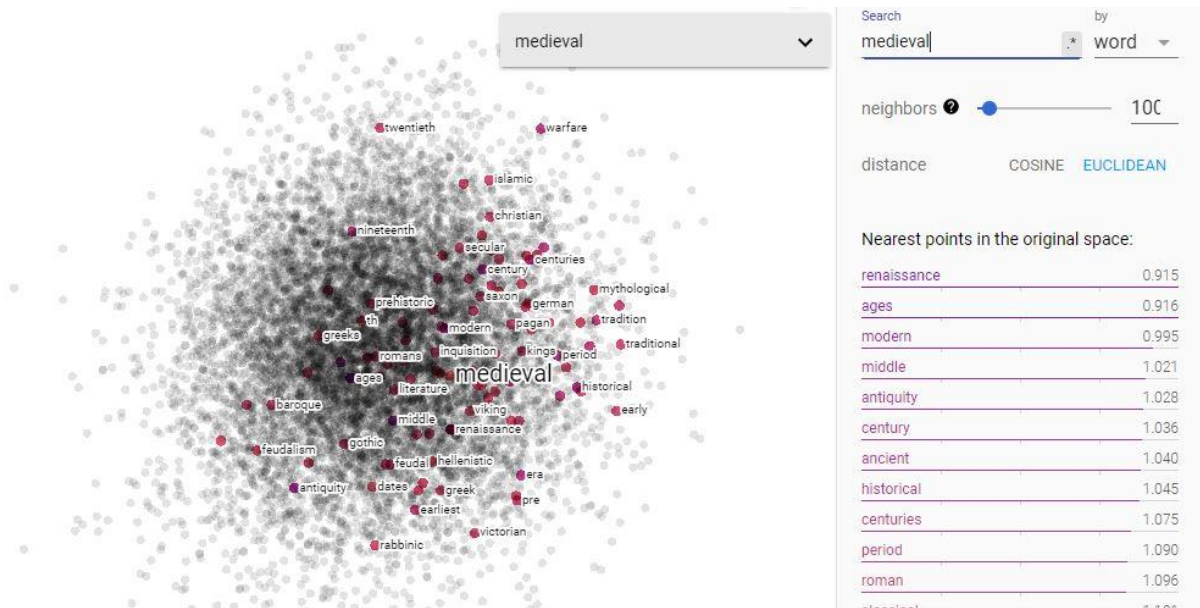
Om het model nog beter te maken, wordt er vaak gebruik gemaakt van nog een andere laag, een *Convolutional Neural Network* (CNN), die voor de (Bi)LSTM-laag wordt geplaatst, waarmee in dit geval een CNN-BiLSTM-model ontstaat.⁴¹ In deze laag ondergaat de tekst al een eerste filtering, waardoor het model uiteindelijk sneller en beter wordt. CNN's zijn origineel ontwikkeld om modellen te kunnen bouwen waarin afbeeldingen de data zijn. Deze laag traint filters, die met een bepaald aantal pixels bijvoorbeeld 4 bij 4, tegelijk bekijkt, en met dat *window* over de afbeelding schuift. De belangrijke pixels worden vervolgens op basis van een toegekende score doorgelaten naar de volgende laag. Deze CNN's bleken ook voor teksten goed te werken, als is nog niet helemaal duidelijk waarom en hoe. Het lijkt er in ieder geval op dat een filter bepaalde typen relevante woorden en woordcombinaties leert herkennen, zoals 'middeleeuwse toestanden'.⁴²

Om een computer met tekst te kunnen laten werken is er ten slotte nog extra stap nodig die de tekst omzet in cijfers. Dit proces heet *word embedding*, en vindt plaats in de eerste laag van een *neural network*, die dan ook de *embeddings*-laag heet.

Het idee achter deze *word embeddings* is dat woorden betekenis hebben in de context en in relatie tot andere woorden. De *embeddings* worden getraind op een grote hoeveelheid teksten, waarbij het leert welke woorden ongeveer hetzelfde betekenen, of welke woorden vaak samen voorkomen. Op basis hiervan worden woorden in een soms wel 300 dimensionale ruimte geplaatst, en ontstaan clusters van woorden die ongeveer een zelfde betekenis hebben, of die in een bepaalde categorie zitten. Het zijn vervolgens de coördinaten van de woorden waarmee het model gaat rekenen. Een voorbeeld van versimpelde *word embeddings* in drie dimensies, is te zien in figuur 1.

⁴¹ Rhanoui, Maryem e.a., 'A CNN-BiLSTM Model for Document-Level Sentiment Analysis', *Machine Learning and Knowledge Extraction* 1 (2019) 832–847.

⁴² Alon Jacovi, Oren Sar Shalom, en Yoav Goldberg, 'Understanding Convolutional Neural Networks for Text Classification', in: *Proceedings of the 2018 EMNLP Workshop Blackbox NLP: Analyzing and Interpreting Neural Networks for NLP* (Brussel, 2018) 56–65.



Figuur 1. Voorbeeld van Engelse *word embeddings* in drie dimensies. De woorden die het dichtst bij 'medieval' staan zijn geselecteerd. Bron: Tensorflow, 'Embedding projector. Visualization of high-dimensional data', <https://projector.tensorflow.org/> (04-07-2020).

Omdat er voor het trainen van *word embeddings* heel veel data en tijd nodig is, heb ik voor dit onderzoek gebruik gemaakt van al getrainde *embeddings*. Hierin zitten meer dan een miljoen woorden, waarin niet alleen perfect Nederlandse woorden zijn opgenomen, maar ook dialecten, waardoor er ook veel negentiende-eeuwse woorden, en/of spellingsvarianten in zitten.⁴³ Uit deze *embeddings* heb ik uiteindelijk de in de ruim 61.000 meest voorkomende woorden uit de trainingsdata gebruikt.

Na verschillende experimenten met verschillende instellingen en samenstellingen, bleek het model met de lagen zoals in figuur 2 het best te werken. In appendix 'B Model' staat een uitgebreidere toelichting op het model, met daarbij ook de belangrijkste instellingen die ik heb gekozen.

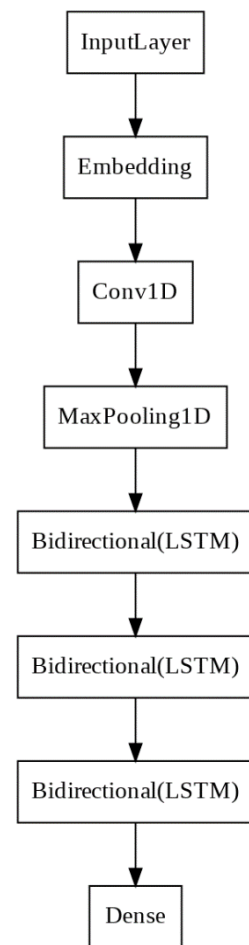
⁴³Stéphan Tulkens, Chris Emmery en Walter Daelemans, 'Evaluating unsupervised Dutch Word embeddings as a linguistic resource', *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016* (2016) 4130–4136.

4.3. De trainingsdata

De trainingsdata voor dit onderzoek bestaat uit 1524 artikelen. Deze artikelen heb ik gelabeld in de categorieën positief (265), negatief (376) of neutraal (540), op basis van de houding ten opzichte van de middeleeuwen die hierin doorscheen. Deze 1524 artikelen zijn gelijkmatige over de periode 1840-1900 verspreid.

In eerste instantie had ik nog voor ik aan deze scriptie begon bij wijze van test zelf zo'n 300 artikelen gelabeld zonder duidelijke, geschreven, richtlijnen. Vervolgens heb ik daarom, en voor de herhaalbaarheid en transparantie, een codelijst opgesteld met de criteria voor ieder van de verschillende categorieën: negatief, neutraal en positief. Dit is ook de gebruikelijke werkwijze bij een handmatige inhoudsanalyse.⁴⁴ Deze codelijst is terug te vinden in de appendices, als appendix 'A. Codelijst'. Toch bleef het nog lastig om het verschil tussen neutraal en positief scherp te houden wat daarmee ook lastig werd voor het model, waardoor ik heb moeten besluiten de neutrale en positieve artikelen als één niet-negatieve categorie te zien. In de volgende paragraaf waarin ik uitleg over de prestaties van het model ga ik hier nog nader op in.

De artikelen heb ik voordat ik ze gebruikte om te trainen op een aantal punten bewerkt, om ze beter geschikt te maken voor het model. Ik heb ervoor gekozen om ieder artikel in te korten tot 250 woorden, waarbij het woord 'middeleeuwen' in het midden staat, voor zover dat mogelijk was. Deze hoeveelheid woorden bleek in de praktijk voor mijzelf meer dan voldoende om het gebruik van het woord te beoordelen, waardoor ook een model hieraan genoeg zou moeten hebben. Meer tekst maakt het voor een algoritme lastiger om te leren, omdat die van te voren niet weet dat het gaat om dat wat er over de middeleeuwen wordt gezegd. Bovendien komt het regelmatig voor dat in met name kranten van vóór 1870, Delpher meerdere artikelen als één artikel worden gezien, waardoor er voor een model heel veel irrelevante en verwarrende tekst in zo'n artikel zit. Overigens ontdekte ik dat de losse artikelen binnen wat volgens Delpher één artikel zag, vaak werden gescheiden



Figuur 2. Schematische weergave van het model.

⁴⁴Neuendorf, *The content analysis guidebook*, 222.

door wat de OCR zag als een langer liggend streepje. Door artikelen op die punten te scheiden, heb ik uiteindelijk veel van dat soort artikelen toch kunnen herstellen.

Daarnaast heb ik zoals gebruikelijk is de leestekens verwijderd en hoofdletters omgezet in kleine letters, hoewel ik de zinseindes wel heb gemarkeerd. Tijdens het labelen van de data ontdekte ik een aantal veelvoorkomende OCR-fouten, die ik zoveel mogelijk heb geprobeerd te herstellen, zoals 'cl' waar eigenlijk een 'd' moet staan. Ook heb ik woorden die vanwege de zeldzame spellingswijze niet in de *embeddings*-laag zitten aangepast naar een spellingswijze die er wél in zit, zoals woorden waarbij een 'y' is gebruikt, in plaats van een 'ij' of zelfs 'ie'.

4.4. Prestaties

4.4.1. Beperkingen van het model

Zoals ik hierboven schreef, heb ik na het trainen van het model op alle drie de categorieën, moeten besluiten om de neutrale en positieve categorieën samen te voegen tot één niet-negatieve categorie. Al tijdens het trainen bleek het erg lastig om deze twee categorieën van elkaar te onderscheiden. Aanvankelijk dacht ik dat dit kwam door een te kort aan positieve voorbeelden, waardoor ik geregeld nieuwe artikelen aan de trainingsdata heb toegevoegd. Uit een nadere analyse bleek wat het probleem was: de lijn tussen neutraal en positief was te dun, of in ieder geval, te vloeibaar. Hierdoor had ik soms vergelijkbare artikelen zowel als positief als neutraal had gelabeld. Ondanks dat ik een codelijst had gemaakt met algemene richtlijnen, bleek dat er toch nog veel lastig te classificeren artikelen waren.

Uiteindelijk realiseerde ik mij dat er een andere meer fundamentele oorzaak voor de fluïde grenzen lag in de gekozen categorieën zelf: negatief, neutraal en positief. Of iemand negatief over de middeleeuwen denkt, is vaak vrij snel duidelijk, maar voor de positieve categorie was dat veel lastiger. Iemand die negatief over de middeleeuwen denkt wil daar niet naar terug, en vind alles in die tijd slecht. Maar zou iemand die positief over de middeleeuwen denkt daar dan wél naar terug willen, en vind zo iemand dan alles in die tijd goed en mooi? En als iemand die positief over de middeleeuwen denkt genuanceerd mag zijn, wanneer is het dan zo genuanceerd dat diegene in de neutrale categorie valt?

Ik kwam tot de conclusie dat daarom een betere indeling zou zijn: desinteresse/negatief, neutraal/overig, interesse/positief. Dit lijkt inderdaad erg op de indeling die ik heb gebruikt, maar is anders omdat bijvoorbeeld iemand die genuanceerd is duidelijk in de interesse categorie valt, evenals een artikel waarin de historische achtergrond van een bepaald

fenomeen wordt geschetst, zonder dat hieraan een duidelijk waardeoordeel wordt verbonden. In een volgend soortgelijk onderzoek is het dus raadzaam voor een dergelijke indeling te kiezen. Daarnaast is het een goed idee om ook een tussen-, of twijfel-categorie te maken voor zowel negatief-neutraal en neutraal-positief. Het voordeel van zo'n categorie is dat ze eventueel bij het trainen van een model er uit gelaten kunnen worden, of dat ze minder zwaar meegewogen kunnen worden in het trainings- en/of evaluatieproces van een model. Maar, vooral helpt het om tijdens het labelen om de andere categorieën zuiver te houden.

Een andere manier om het probleem te voorkomen waarmee ik nu te maken heb, is om tijdens het labelen voor ieder artikel in de trainingsdata, ook het onderwerp te coderen van dat waarbij de middeleeuwen worden betrokken. Dit is nuttig om twee redenen. Allereerst kunnen deze onderwerpen gebruikt kunnen worden voor het trainen van een model die de artikelen ook op onderwerpen classificeert, waardoor het bijvoorbeeld makkelijker wordt om te verklaren waarom een bepaalde groep positiever of negatiever over de middeleeuwen schrijft, of waarin de interesse verschilt tussen bepaalde groepen. Daarnaast kunnen deze onderwerpen helpen om achteraf fouten in het classificeren op interesse te herstellen. Ik had bijvoorbeeld aanvankelijk bedacht om verslagen over parades en feesten met middeleeuwse elementen op te nemen in de positieve categorie, en lezingen, als ze niet duidelijk positief waren, als neutraal; achteraf had ik dit graag anders willen doen. Met een model dat artikelen ook onderwerp classificeert zou ik dat achteraf gemakkelijk kunnen doen. Een andere manier om hetzelfde te bereiken, zou kunnen door simpelweg de trainingsdata aan te passen door te besluiten artikelen met een bepaald onderwerp in een andere houding-categorie in te delen, en vervolgens het classificatiemodel opnieuw te trainen.

Helaas had ik geen tijd meer om deze geleerde lessen toe te passen in dit onderzoek, waardoor er nu dus twee categorieën overblijven.

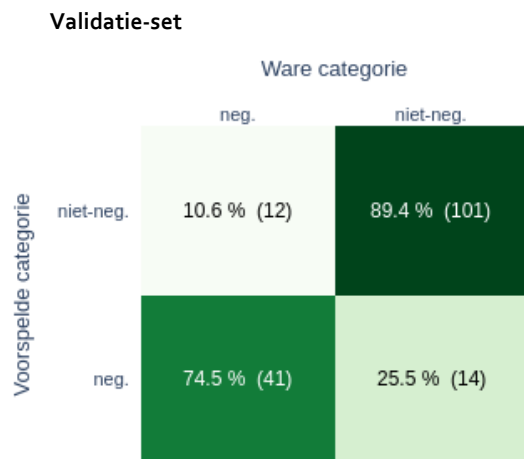
4.4.2. Prestatiesstatistieken

De negatieve en niet-negatieve categorieën worden door het model wél goed van elkaar onderscheiden. Dat brengt ons bij de prestatiestatistieken van het model. Om duidelijk te maken waarop de betrouwbaarheid bepaald wordt, is het nodig om eerst nog iets over het trainingsproces uit te leggen.

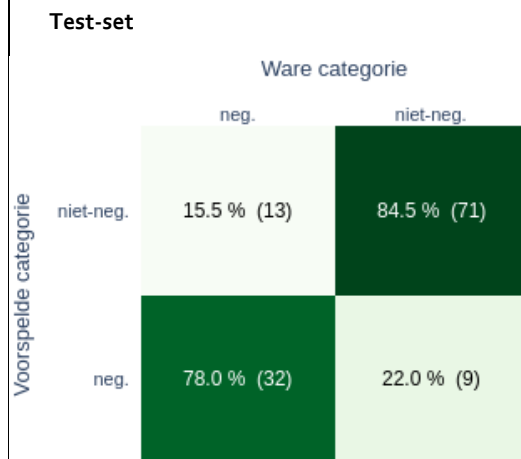
In het trainingsproces wordt de trainingsdata verdeeld in drie groepen: een training-set, een validatie-set en een test-set, waarbij de training-set meestal zo'n 80% van de trainingsdata

bevat. Tijdens het trainen wordt de training-set gebruikt om het model daadwerkelijk te trainen, en dus de parameters aan te passen. Nadat alle trainingsvoorbeelden door het model zijn gegaan, gaat de validatie-set door het model. Maar, bij de validatie-set wordt het model niet aangepast. Hierdoor kan tijdens het trainen in de gaten worden gehouden of het model ook daadwerkelijk algemene regels leert om een artikel te kunnen classificeren, of dat het slechts de artikelen uit de training-set leert te herkennen die bij een bepaalde categorie horen. In dat laatste geval is er sprake van *overfitting*. De bedoeling is dat de training-set en de validatie-set min of meer even goed worden geclassificeerd. Als dat uiteindelijk het geval is, en het model goed genoeg werkt, is er nog de test-set met voorbeelden die het model helemaal niet heeft gezien. Als ook hierop het model ongeveer even goed scoort als op de validatie-set, dan kan er vanuit gegaan worden dat het model in de praktijk min of meer net zo goed werkt. Daarom zal ik hier van zowel de validatie- als de training-set de statistieken geven.

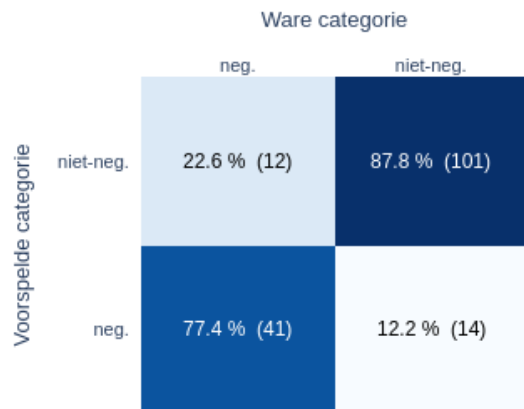
De accuraatheid (gedeelte wat juist is geclassificeerd) op de test-set is 82,4% en op de validatie-set 84,5%. De test-set en validatie-set samen komen uit op 83,6%. In figuur 3 staan voor de beide sets twee zogenaamde *confusion matrices*, met gedetailleerder prestaties. Hierin zijn de voorspelde en ware labels van alle voorbeelden samengevat, waarbij het kwadrant rechtsboven bijvoorbeeld aangeeft hoeveel van de niet-negatieve voorbeelden ook daadwerkelijk als niet-negatief zijn geclassificeerd, en in het kwadrant linksboven af te lezen is hoeveel niet-negatieven er (foutief) als negatief zijn beoordeeld. Tussen achter de percentages staat tussen haakjes het absolute aantal voorbeelden voor dat kwadrant. In 3a en 3b is met een percentage aangegeven welk deel van een categorie ook als die categorie is beoordeeld, en in c en d is de zuiverheid van categorieën aangegeven, dat wil zeggen, welk deel van wat in een bepaalde categorie is geclassificeerd, ook daadwerkelijk tot die categorie behoort.



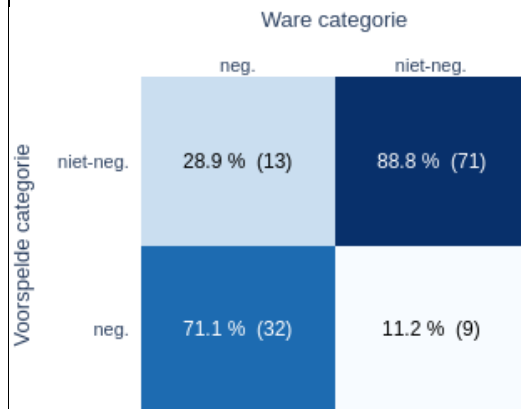
Figuur 3a. Confusion matrix validatie-set percentage van categorie juist of niet juist geïdentificeerd.



Figuur 3b. Confusion matrix test-set percentage van categorie juist of niet juist geïdentificeerd.



Figuur 3c. Confusion matrix validatie-set percentage van categorie wat ook tot die categorie behoort of niet.



Figuur 3d. Confusion matrix test-set percentage van categorie wat ook tot die categorie behoort of niet.

4. Conclusie

In dit eerste deel stond de vraag centraal hoe *neural networks* gebruikt kunnen worden door historici in een onderzoek naar hoe de ontwikkeling in de houding van een samenleving ten opzichte van een bepaald onderwerp, in dit geval de middeleeuwen, en hoe dit is te rechtvaardigen. Daarbij hebben we gezien dat deze methode een volgende stap zou kunnen zijn binnen de historiografie en de *digital humanities*, waarbij historici gebruik kunnen maken van de alsmaar groeiende hoeveelheid gedigitaliseerde bronnen. Nu de technologische beperkingen voor het gebruik van geavanceerde technieken als *neural networks* in rap tempo afnemen, is het ook voor historici zelf mogelijk om deze technologie te leren en vervolgens ook toe te passen.

Het is daarnaast goed om ook de beperkingen van deze methode te erkennen, namelijk dat ook deze manier van onderzoek doen een andere kijk op het verleden biedt, en daarin eerder een aanvulling dan een vervanging is van traditioneel historisch bronnen onderzoek. De data die hiermee gecreëerd wordt kan bijvoorbeeld gebruikt worden als startpunt van verder kwalitatief onderzoek, die verder ingaat op eventuele opmerkelijke patronen in die data.

Tot slot heb ik laten zien hoe het CNN-BiLSTM-model wat ik in dit onderzoek gebruik er uit ziet, hoe het tot stand gekomen is en hoe het werkt. Achteraf bleek echter dat de neutrale en positieve categorieën niet goed genoeg afgebakend waren, wat deels het gevolg was van de keuze voor deze categorieën überhaupt. Hierdoor zijn er nu twee categorieën: negatief en niet-negatief.

Desalniettemin gaat dit niet ten koste van het onderzoek, vanwege de inzichten die de fouten hebben opgeleverd enerzijds, en anderzijds blijft ook met twee categorieën het doel van dit onderzoek binnen bereik: het onderzoeken of de toepassing van deze technieken meer inzicht kan geven in de ontwikkeling van de houding ten opzichte van de middeleeuwen.

Deel Twee: Resultaten

5. Introductie

Na het meer theoretische eerste deel, zal ik hier in dit meer toegepaste tweede deel, de data die ik met het model heb geproduceerd presenteren en analyseren, om daarmee antwoord proberen te geven op de vraag hoe de houding ten opzichte van de middeleeuwen in Nederland in negentiende eeuw ontwikkelde. Zoals ik in de algemene inleiding al deels besprak, is het mooie aan het gebruik van een grote hoeveelheid krantenartikelen al bron, dat het data oplevert die uitgesplitst kan worden naar waarnaar de kranten zelf zijn uit te splitsen, zoals sociale klasse, ideologie, religie, stad, regio en provincie. Hieraan zit voor deze scriptie waarin ook de methode een grote plek inneemt, in zekere zin een nadeel, omdat er geen ruimte is om de data op alle verschillende niveaus en in alle combinaties te analyseren. In dit deel zal ik mij daarom beperken tot de ontwikkeling van verschillende zuilen of ideologische stromingen op nationaal niveau, omdat ik daarmee bovendien kan aansluiten op eerder onderzoek.

Allereerst zal ik de ruwe niet-geclassificeerde data beschrijven, zodat onder andere duidelijk wordt hoeveel artikelen en kranten ik heb gebruikt per zuil per vijf jaar. Vervolgens zal ik de geclassificeerde ruwe data beschrijven en analyseren, waarbij ik de meest opvallende ontwikkelingen aanstip. Hierna zal ik de ontwikkelingen in met name de katholieke en protestantse zuil gedeeltelijk proberen te verklaren. Dit doe ik door de grafieken te combineren met een tijdlijn met daarop publicaties die volgens Peter Raedts invloedrijk waren voor de perceptie van de middeleeuwen in de negentiende eeuw. Zo kunnen we zien of, en in welke mate deze publicaties ook daadwerkelijk invloed hadden op de houding ten opzichte van de middeleeuwen zoals die zichtbaar is in krantenartikelen.

6. Ruwe data

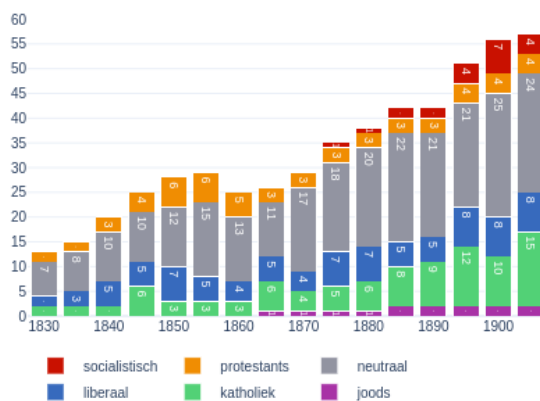
Voordat ik per zuil de resultaten zal beschrijven en analyseren, is het belangrijk eerst te kijken naar de gebruikte data, waaruit we mogelijk al een aantal inzichten kunnen destilleren. Ik heb gebruik gemaakt van alle krantenartikelen waarin de middeleeuwen genoemd worden die tussen 1830 en 1910 zijn verschenen in kranten die binnen Nederland zijn verspreid, en die op 23 april 2020 via Delpher beschikbaar waren. Dit zijn helaas niet alle gedigitaliseerde Nederlandse krantenartikelen uit deze periode. Ook andere regionale en provinciale archieven hebben kranten gedigitaliseerd, maar die hebben allemaal hun eigen platforms waarop deze kranten te bekijken zijn, en bieden daar niet een optie om ze massaal te downloaden zoals de Koninklijke Bibliotheek die wel biedt. Via Delpher zijn enkele van die archieven overigens inmiddels wel te doorzoeken, en dit worden er in de toekomst meer, maar artikelen zijn (vooralsnog) niet te via die weg te downloaden.⁴⁵ Vanwege de beperkte tijd voor dit onderzoek, heb ik mij daarom nu alleen gericht op de artikelen die Delpher staan.

Desalniettemin blijft er een grote hoeveelheid artikelen en kranten over die wel meegenomen worden. In tabel 1 is het aantal artikelen per vijf jaar, per zuil weergegeven.

⁴⁵ Delpher, 'Wat zit er in Delpher?' <https://www.delpher.nl/nl/platform/pages/helpitems?nid=395> (geraadpleegd op 6 juli 2020).

Jaren	joods	katholiek	liberaal	neutraal	protestants	socialistisch	totaal
1830	0	32	52	77	38	0	199
1835	0	27	68	133	41	0	269
1840	0	41	127	104	28	0	300
1845	0	106	205	149	48	0	508
1850	0	116	110	89	71	0	386
1855	0	96	111	171	65	0	443
1860	0	114	124	144	51	0	433
1865	31	111	216	186	27	0	571
1870	2	179	529	401	44	0	1155
1875	19	284	577	618	51	4	1553
1880	22	335	581	585	97	14	1634
1885	35	548	452	763	192	84	2074
1890	68	781	630	1060	270	163	2972
1895	98	1030	1056	1714	362	354	4614
1900	82	1246	988	2037	442	351	5146
1905	92	1712	1661	2362	481	317	6625
Totaal:	449	6758	7487	10593	2308	1287	28882

Tabel 1. Totaal artikelen per zuil per vijf jaar.



Figuur 4. Aantal unieke kranten per zuil per vijf jaar.

zuil	aantal
joods	2
katholiek	23
liberaal	20
neutraal	64
protestants	15
socialistisch	8

Tabel 2. Totaal beschikbare kranten per zuil in de periode 1830-1909.

Hierin valt meteen op dat er geen joodse en socialistische artikelen zijn tot respectievelijk 1865 en 1875. Wat betreft de socialistische artikelen is de verklaring eenvoudig, namelijk dat het socialisme pas rond 1875 begon te groeien in Nederland, en in 1881 de eerste socialistische partij (de Sociaal-Democratische Bond) werd opgericht.⁴⁶ Voor de joodse artikelen is de verklaring soortgelijk, in die zin, dat pas in 1865 de eerste, en voor twintig jaar de enige, joodse krant werd opgericht: het *Nieuw Israëlitisch Nieuwsblad*.⁴⁷ De tweede joodse

⁴⁶ Friso Wielenga, *De geschiedenis van Nederland. Van de Opstand tot heden* (Amsterdam 2012) 250.

⁴⁷ Jaap Meijer, 'Arduous adaptation, 1814-1870', in: J. C. H.; Blom, R. G.; Fuks-Mansfeld en I. Schoffer ed., *History of the Jews in the Netherlands* (Portland, Oregon 2001) 192-229, aldaar 222.

krant die binnen de hier onderzochte periode valt, werd in 1885 opgericht: het *Centraalblad voor Israëlieten in Nederland*.⁴⁸ Andere joodse kranten zijn er voor deze periode niet.

Dit brengt ons bij figuur 4, waarin het aantal unieke krantentitels dat gebruikt is in dit onderzoek per zuil per vijf jaar is weergegeven. Het overgrote deel van de kranten is neutraal (64), dat wil zeggen, niet uitgesproken aan een specifieke zuil gebonden. Het aantal katholieke (23) en liberale (20) kranten is ongeveer gelijk, met iets daaronder de protestantse kranten (15). In figuur 4 is per vijf jaar het aantal unieke kranten per zuil weergegeven. Hierin valt op dat de verhoudingen tussen de aantallen kranten per zuil voor het grootste deel gedurende deze periode min of meer gelijk is gebleven. Een uitzondering is de groep protestantse kranten, waarvan de grootte weinig veranderd, en dus relatief een steeds kleiner aandeel van het geheel vormt.

Ook valt in zowel tabel 1 als figuur 4 op, dat er een behoorlijke toename is van het aantal kranten en artikelen vanaf 1870. Deze toename is voornamelijk te wijten aan de afschaffing van de dagbladzegel. Deze zegel was een soort accijns op kranten.⁴⁹ De afschaffing hiervan betekende dus dat kranten veel goedkoper werden. Ook was het een drempel om te voorkomen dat iedereen zomaar een krant kon beginnen.⁵⁰ Na afschaffing van de zegel groeide het aantal kranten dan ook fors. Daarnaast versnelde vanaf deze tijd de groei van nieuwe genres in kranten; een proces wat vanaf 1830 was begonnen.⁵¹ Naast de traditionele nieuwsberichten verschenen er steeds meer opiniestukken, vervolgh verhalen, gedichten en reportages (zie figuur 7, waarbij deze laatste drie genres onder 'overig' vallen).

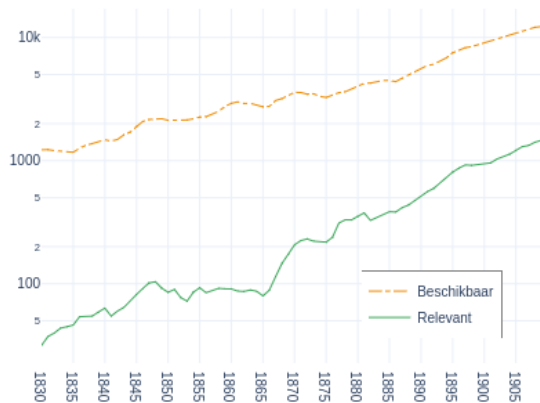
Desalniettemin zou het kunnen dat het aantal artikelen waarin over de middeleeuwen wordt gesproken harder of minder had groeit dan de groei van het totale aantal artikelen. Later zal ik hier nog wat verder op ingaan, als ik de resultaten per zuil ga vergelijken. Voor een algemeen beeld heb ik in figuur 5 zowel het totaal aantal beschikbare artikelen op Delpher in deze periode weergegeven, als het totaal aantal relevante artikelen voor dit onderzoek in dezelfde periode (totalen per vijf jaar). De beschikbare artikelen zijn alle gedigitaliseerde artikelen van de kranten die ik in dit onderzoek gebruik. In de grafiek zijn de

⁴⁸ J. C. H. Blom en J.J. Cahen, 'Jewish Netherlanders. Netherlands Jews, and Jews in the Netherlands, 1870–1940', in: J. C. H.; Blom, R. G.; Fuks-Mansfeld en I. Schoffer ed., *History of the Jews in the Netherlands* (Portland, Oregon 2001) 230–295, aldaar 259.

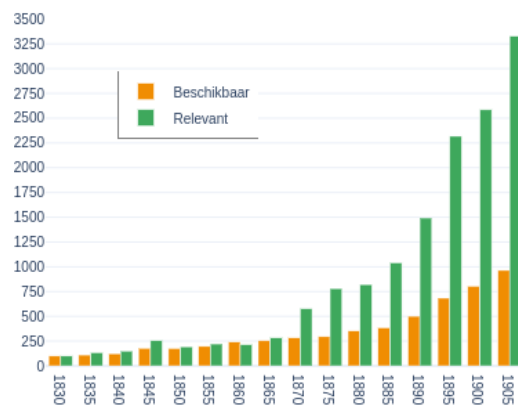
⁴⁹ De Graaf, 'Journalistiek in beweging', 34-35.

⁵⁰ Ibidem.

⁵¹ Idem, 174-175.



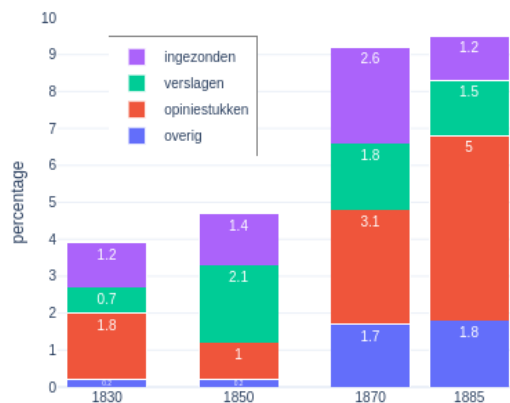
Figuur 5. Beschikbare artikelen op Delpher van kranten waaruit artikelen zijn genomen, versus middeleeuwen artikelen uit die kranten. De aantallen zijn op logaritmische schaal, waardoor de relatieve groei beter te vergelijken is. De kleine getallen op de y-as stellen zijn niet de logaritmische waarden, maar moeten vermenigvuldigd worden met het 100, 1000 of 10000-tal waarin ze ligt. De eerste 2 op de y-as staat daarmee voor 200 (2*100), de tweede 2 voor 2000 (2*1000).



Figuur 6. Index van de beschikbare, en relevante artikelen op Delpher. Het startpunt is het aantal artikelen in 1830. De index is als volgt berekend:

$$\text{index} = (\text{waarde [jaar]} / \text{waarde 1830}) * 100.$$

absolute aantallen op een logaritmische schaal weergegeven, zodat de relatieve groei beter te vergelijken is en ook de kleinere aantallen beter zichtbaar worden. Om de relatieve groei nog beter te kunnen vergelijken, is in figuur 6 dezelfde data weergegeven als een index, waarbij het eerste jaar (1830) als basisjaar is genomen.



Figuur 7. Aandeel genres in kranten in procent. Weggelaten genres zijn : nieuwsberichten, serviceberichten, en overheidsmededelingen. Bron: R.H. de Graaf, 'Journalistiek in beweging. Veranderende berichtgeving in kranten en pamfletten (Groningen en 's-Hertogenbosch 1813-1899)', (Proefschrift Utrecht 2009), 104-115.

In de absolute aantallen is inderdaad te zien dat het aantal beschikbare artikelen stijgt hard stijgen vanaf 1870. Maar in figuur 6 valt onmiddellijk op dat het aantal relevante artikelen veel sneller en harder stijgt dan het aantal beschikbare.

In het eerste deel van de grafiek valt op dat er tot 1849 een gestage groei is van zowel het aantal beschikbare als het aantal relevante artikelen, waarbij het aantal relevante artikelen consequent harder groeit dan het aantal beschikbare. Na deze piek van 255 punten van de relevante artikelen (2,25 keer meer artikelen dan in 1830), daalt dit weer opvallend hard naar

193 punten, en duurt het tot de 1864 voordat dit weer is ingehaald (286 punten). Tegelijkertijd daalt het aantal beschikbare artikelen met maar één punt naar 177.

In de jaren '50 groeien de relevante en beschikbare artikelen weer gelijk op, totdat er vanaf 1860 grote verschillen ontstaan. Vanaf 1860 groeit het aantal relevante artikelen tot 1879 met 72%, terwijl het aantal beschikbare artikelen met maar 16% groeit in dezelfde periode. Vanaf 1880 stijgen beide weer behoorlijk, de relevante artikelen met 75%, en de beschikbare met 63%.

De hardere groei van het aantal relevante artikelen kan tenminste voor een groot deel verklaard worden door de groei van de nieuwe genres in deze periode. In deze nieuwe genres werd er waarschijnlijk meer gesproken over de middeleeuwen dan in de traditionele nieuwsberichten, hoewel het ook waarschijnlijk is dat de middeleeuwen überhaupt populairder werden. Overigens hoeft die populariteit niet noodzakelijk ook te liggen in Nederland, en hoeft het ook niet zo te zijn dat die populariteit ook niet-negatief was (in Nederland). Een krant publiceert immers niet alleen wat er in het eigen land gebeurt, maar ook daar buiten. Ceremonies of festiviteiten in het buitenland waarin middeleeuwse elementen zitten, zoals kostuums, gebruiken of spelen, werden ook vermeld. Hetzelfde geldt voor 'middeleeuwse toestanden', zoals pogroms die in onder andere Duitsland en Oostenrijk nog wel eens voorkwamen.

Over de vraag in hoeverre er negatief en niet-negatief geschreven werd over de middeleeuwen als ik mij in het volgende hoofdstuk buigen, waarin ik zal ik de data van de geclassificeerde artikelen zal analyseren.

7. De middeleeuwen per zuil

In dit hoofdstuk zal ik antwoord proberen te geven op de vraag hoe de houding per zuil ten opzichte van de middeleeuwen ontwikkelde in de periode 1830-1910. Omdat er maar twee van de bedoelde drie houdingen zijn overgebleven om te analyseren, is het analyseren van enkel de negatieve houding voldoende om ook inzicht te krijgen in de tegenovergestelde niet-negatieve houding. Nadat ik eerst de zal beschrijven hoe deze houding ontwikkeld, wat daarin opvalt, en waar er (grote) verschillen zijn tussen zuilen onderling, zal ik daarna testen of deze ontwikkelingen en verschillen te verklaren zijn aan de hand van enkele publicaties in de negentiende eeuw, die volgens Peter Raedts belangrijk zijn geweest in de beeldvorming van de middeleeuwen.

7.1. Data en analyse

De ontwikkeling van de houding tot middeleeuwen gedurende de negentiende eeuw, blijkt een opmerkelijke te zijn. In figuur 8 is de 'negativiteitsratio', dat wil zeggen de verhouding van het aantal negatieve ten opzichte van het aantal niet-negatieve artikelen, voor de verschillende zuilen weergegeven. Om vooral inzicht te krijgen in de trend over de langere termijn, en niet zo zeer in jaarlijkse uitschieters, zijn deze gegevens weergegeven als een voortschrijdend gemiddelde van vijf jaar. Dat betekent dat voor ieder jaar is gemiddeld met de twee jaar ervoor en erna. Ook heb ik het algemene gemiddelde, of de algemene trend weergegeven. Dit gemiddelde is berekend over alle zuilen, met uitzondering van de joodse en

socialistische zuil vanwege het vertekende beeld wat dit zou opleveren. Bovendien is tot het verschijnen van socialistische en joodse kranten, het gemiddelde ook steeds genomen over de overige zuilen. Om ook inzicht te krijgen in de verschillen tussen de zuilen zelf, zijn in figuur 9 dezelfde gegevens minus het algemene gemiddelde weergegeven.

Om te weten hoe groot de kans is dat bepaalde verschillen toevallig zijn, heb ik voor iedere vijf jaar (voortschrijdend) een Kruskal-Wallis-test gedaan. Ik maak gebruik van deze test omdat de data niet voldoet aan de voorwaarden voor de standaard analyse, de One-Way ANOVA.⁵² In een Kruskal-Wallis-test worden de observaties op basis van hun waarden (de negativiteitsratio) gerangschikt, waarna met de betreffende rangnummers voor iedere zuil een statistiek wordt berekend.⁵³ Met de resultaten van deze test kan iets gezegd worden over in hoeverre het verschil tussen de groepen toevallig of willekeurig is. Naast een teststatistiek wordt er ook een p-waarde berekend. Als die waarde lager is dan een bepaalde grenswaarde (*alpha*), dan is het verschil significant. Hier heb ik gekozen voor een *alpha* van 0.1, onder meer omdat de datasets voor de vijf jaren vrij klein is.⁵⁴ Dit betekent dat wanneer een verschil significant is, de kans dat dit verschil aan het toeval is toe te schrijven kleiner is dan tien procent. De p-waardes heb ik een *heatmap* weergegeven in figuur 10 waarbij de groene vakjes een significant resultaat aangeven. Ik heb voor deze weergave gekozen omdat het niet mogelijk is om alle testgegevens in tabellen weer te geven zoals dat gebruikelijk is; voor iedere combinatie heb ik deze test immers tachtig keer deze uitgevoerd.

Als input voor de test gebruik ik de negativiteitsratio per krant voor ieder jaar in de vijf die geanalyseerd wordt. Iedere krant wordt zo dus voor ieder jaar een observatie voor de zuil waartoe deze behoort. Omdat in de grafiek alle kranten voor een zuil voor een jaar wel tot één waarde zijn gereduceerd, kan het zijn dat er grote verschillen zitten binnen een zuil, waardoor er een verschil lijkt te zijn tussen zuilen wat er misschien niet per se is. Deze test voorkomt zo dus dat ik onterechte conclusies trek op basis van de grafiek. In de test nam ik voor een jaar overigens alleen een krant mee als er in die krant voor dat jaar zeven of meer relevante artikelen stonden.

Naast dat ik per zuil heb getest, heb ik ook getest op verschil tussen confessionele (protestantse en katholieke) en niet-confessionele kranten (de overige zuilen), en tussen

⁵² B Blondé e.a., *Trend en toeval. Inleiding tot de kwantitatieve methoden voor historici*. (Leuven 2015) 247,264.

⁵³ Ibidem.

⁵⁴ Andrew Noymer, 'Alpha, Significance Level of Test', *Encyclopedia of Survey Research Methods* (Thousand Oaks 2011) 17–19.

liberale en neutrale, en overige kranten. Dit laatste heb ik gedaan omdat de liberale en neutrale groep soms als één (niet-)zuil wordt gezien. Daarnaast gaven de grafieken, evenals voor het vergelijken van de confessionelen met de niet-confessionelen, aanleiding om dit zo te testen.

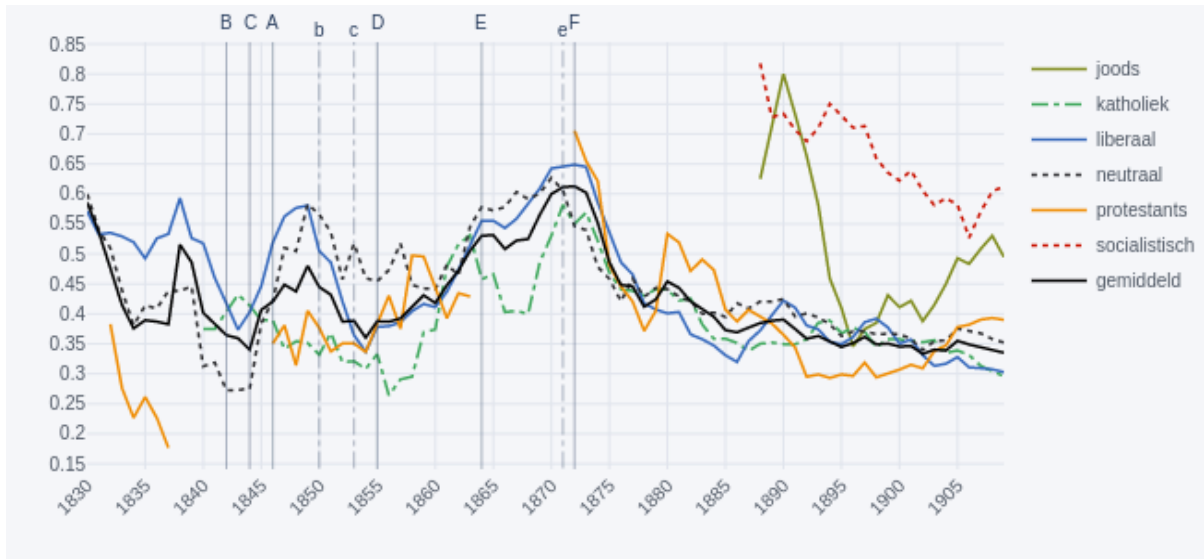
Aangezien de Kruskal-Wallis-test enkel aangeeft óf er een significant verschil zit tussen de groepen, maar niet tussen welke groepen dit verschil significant is, moet er ook een vervolgstap gedaan worden. Hiervoor heb ik de Mann-Whitney-test gebruikt.⁵⁵ Hierin worden steeds twee zuilen met elkaar vergeleken op dezelfde manier als in de Kruskal-Wallis-test: de waarden worden gerangschikt, waarna wordt gekeken of er een verschil zit tussen de plaatsen die de groepen op de rangschikking hebben. Ook hieruit rolt een statistiek en een p-waarde. De p-waarde is hier gecorrigeerd voor de hoeveelheid paren met de Simes-Hochberg-correctie.⁵⁶ Omdat het helemaal onmogelijk is om hiervoor alle tabellen weer te geven, heb ik ook deze p-waarden in een heatmap gezet in figuur 11.

Als we allereerst kijken naar de grafiek in figuur 8, zien we meteen een aantal opmerkelijke ontwikkelingen. In de gemiddelde trend van de van de houding tot de middeleeuwen, valt op dat er drie pieken van negativiteit zijn. De eerste vlak voor 1840, de tweede vlak voor 1850, en de derde vlak rond 1870. Na deze laatste piek neemt de algehele mate van negativiteit in de tien daarop volgende jaren flink af, met een halvering tot ongeveer veertig procent, waarna het vervolgens tot het eind van de periode langzaam verder afneemt richting de dertig procent.

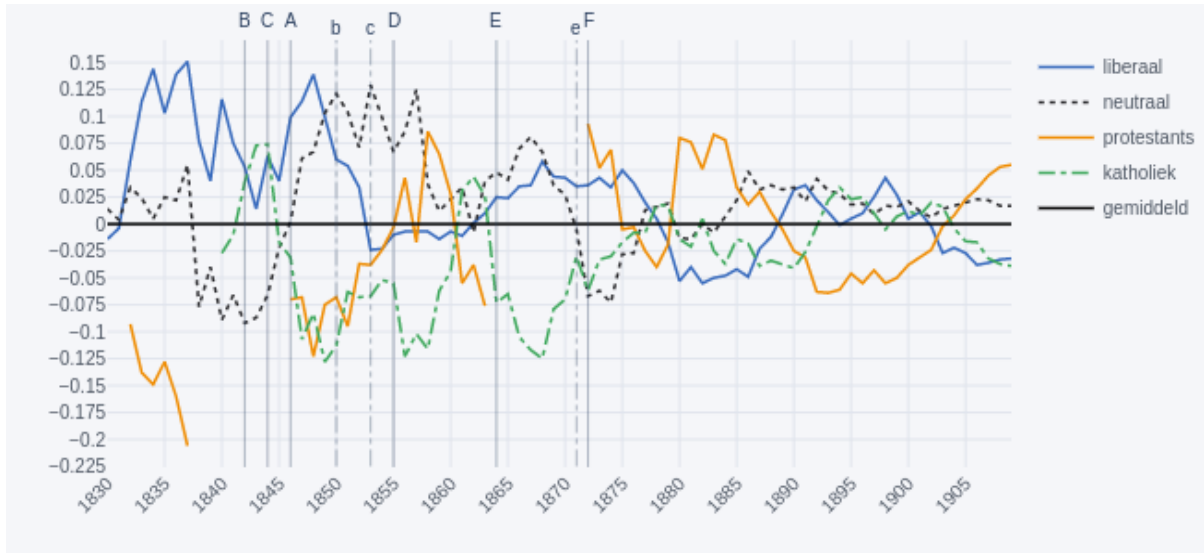
Wat opvalt in de aan pieken, zijn de grote verschillen tussen de verschillende zuilen onderling. Een blik op figuur 9 maakt dit nog duidelijker. Tot 1880 lijken vooral de katholieken tegenover de liberalen en neutralen te staan, en heeft het er bovendien alle schijn van dat wanneer de liberalen negatiever worden, de katholieken juist positiever; hoewel hier wie op wie reageert niet heel duidelijk is. Uit figuur 11 blijkt dat de verschillen tussen deze groepen op die punten in rond de pieken (rond 1848 en 1866) ook significant zijn; zowel tussen die katholieken en liberalen, als tussen katholieken en neutralen ($p < 0.1$).

⁵⁵ Blondé, *Trend en toeval*, 256.

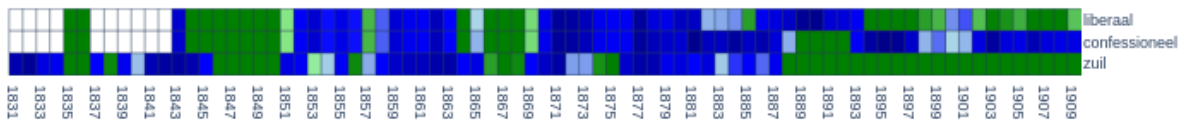
⁵⁶ Yoav Benjamini en Yosef Hochberg, 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1995) 289–300.



Figuur 8. Percentage negatieve artikelen per zuil per jaar (voortschrijdend gemiddelde over vijf jaar). Voor een gegeven jaar is de waarde het gemiddelde van de twee jaar ervoor en erna, en het desbetreffende jaar zelf. Gemiddeld minimaal aantal artikelen per jaar voor de betreffende middel-periode voor een zuil is 10. Vanwege de grote uitschieters van de socialistische en de joodse kranten, zijn die in het algemene gemiddelde niet meegenomen; deze zouden anders een te vertekend beeld geven. De verticale lijnen met letters markeren belangrijke publicaties over de middeleeuwen, waarbij voor publicaties over een langere periode het eindjaar met een kleine letter ook is aangegeven, zie paragraaf 7.2.



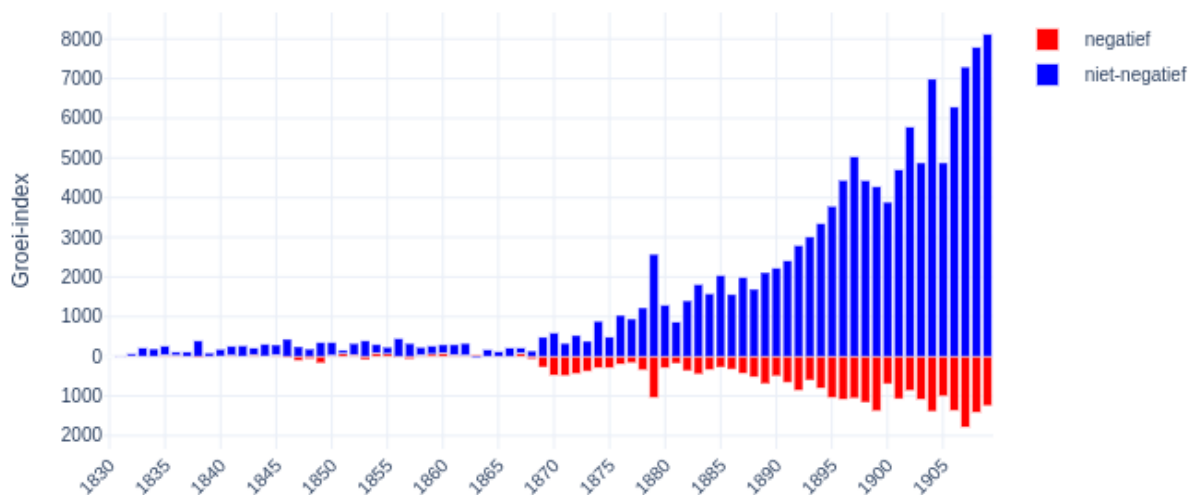
Figuur 9. Afwijking van gemiddeld percentage negatieve artikelen per zuil per jaar (voortschrijdend gemiddelde over vijf jaar). +0.1 betekent hier dus 10 procent meer dan het gemiddelde. Voor details over gemiddelde en uitleg over letter, zie beschrijving 8. De socialistische en joodse kranten zijn weggelaten voor een beter beeld van de grotere zuilen.



Figuur 10. Significantie-heatmap Kruskal-Wallis-test per jaar (voortschrijdend per vijf jaar, twee jaar voor en twee jaar na gegeven jaar). Donkergroen $p \leq 0.05$, lichtgroen $p \leq 0.1$, verloop lichtblauw-blauw $p > 0.1 < 0.2$, blauw-donkerblauw $p > 0.2$. Wit is geen waarde; in deze gevallen was er te weinig data beschikbaar door te weinig groepen, mogelijk veroorzaakt door te weinig data per krant (minder dan 7 artikelen per jaar per krant). Zie tekst voor uitleg over de groepen.



Figuur 11. Significantie-heatmap Mann-Whitney-tests per zuil-paar, liberaal - niet-liberaal, en confessioneel-niet-confessioneel, per jaar (voortschrijdend per vijf jaar, twee jaar voor en twee jaar na gegeven jaar). De laatste twee testen zijn apart getest. De p-waardes zijn gecorrigeerd met de Simes-Hochberg correctie. Donkergroen $p \leq 0.05$, lichtgroen $p \leq 0.1$, verloop lichtblauw-blauw $p > 0.1 < 0.2$, blauw-donkerblauw $p > 0.2$. Wit is geen waarde; in deze gevallen was er te weinig data beschikbaar (minder dan 5 kranten per zuil/groep), mogelijk veroorzaakt door te weinig data per krant (minder dan 7 artikelen per jaar per krant).



Figuur 12. Index negatieve en niet-negatieve artikelen per jaar, gecorrigeerd voor algemene groei (index) van beschikbare artikelen. Waarden zijn de index voor negatieve en niet negatieve artikelen minus de index van de beschikbare artikelen. Startjaar is 1830. Let op dat de waarden van de negatieve artikelen die in het rood onder de nullijn zijn weergegeven, positief zijn.

De afname van negativiteit vanaf 1875, lijkt vooral het gevolg te zijn van een hardere toename van niet-negatieve artikelen. In figuur 12 heb ik de groei-index van de zowel de negatieve als niet-negatieve artikelen berekend, en hiervan vervolgens de groei-index van de beschikbare artikelen afgetrokken. Hierin is te zien dat het aantal niet-negatieve artikelen veel harder groeit dan het aantal negatieve artikelen.

Een grote uitzondering op zowel de mate van negativiteit, als de afnemende trend in die periode, zijn de socialistische en joodse kranten. Tussen 1885-'90 verschijnen zij voor het eerst in de grafiek, op tachtig procent negativiteit. Beiden worden vervolgens met de algemene trend minder negatief, hoewel de negativiteitsratio veel hoger blijft dan de overige zuilen. Vanaf het einde van de jaren '90 wijken de joodse kranten weer af van die trend, en worden ze juist weer veel meer negatiever. Vanaf 1905 laten ook de socialisten weer een stijgende lijn in negativiteit zien. Uit figuur 11 blijken met name de verschillen tussen de verschillen tussen socialisten en de overige zuilen significant zijn, wat ook het geval is voor de joodse kranten rond 1890, en deels ook in de laatste jaren vóór 1910.

Tussen de liberalen en protestanten (en ook katholieken en protestanten) is er tussen 1880-'85 in zowel figuur 9, als in de significantie-heatmap van figuur 11, een opmerkelijk verschil zichtbaar, iets wat overigens ook in 1835-'36, 1848, en 1908 het geval is. Voor de overige jaren blijken de verschillen niet significant. Met name in de periode vanaf 1887 is er tussen neutralen, liberalen protestanten en katholieken geen significant verschil, in de vijf jaren die per jaar worden meegenomen.

Tot slot zijn ook niet significante verschillen een interessant resultaat. Dit laat zien dat er, zoals ook uit de grafieken blijkt, ook periodes waren van weinig echt grote of duidelijke verschillen tussen alle of juist enkele zuilen. Tussen de liberalen en neutralen is er zelfs over de gehele periode geen significant verschil te zien wat overeenkomt met het beeld uit de grafieken. Dit is niet geheel onverwachts, omdat neutralen en liberalen vaak onder één zuil worden geschaard. Bovendien was in de negentiende eeuw het liberalisme in Nederland, in ieder geval in de politiek, en dus in de stemgerechtigde bovenklasse waaruit ook de kranteneigenaars kwamen, de overheersende, en daardoor misschien ook wel de 'neutrale' ideologie.⁵⁷

Overigens betekent een niet significant verschil, of een niet zichtbaar verschil in de grafiek, niet dat er ook per definitie geen verschil tussen bepaalde zuilen bestaat. Soms is er

⁵⁷ Wielenga, De geschiedenis van Nederland, 252.

slechts te weinig data, of lopen de verschillen over andere grenzen dan puur de hoofdzuil waartoe een krant behoort. Het zou goed kunnen dat er bijvoorbeeld wel verschil is tussen progressieve of conservatieve protestanten en katholieken. In deze periode raakte het politieke landschap, en daarmee ook het perslandschap, namelijk veel meer verdeeld en versnipperd.⁵⁸ Helaas is er in deze scriptie geen ruimte om deze eventuele verschillen verder te onderzoeken.

7.2. Verklaringen

Nu de verschillen en overeenkomsten zijn opgemerkt, zou is de volgende stap dat deze verschillen en overeenkomsten geduid worden. Gezien echter de lengte van de periode, en de vele aanknopingspunten die er nu al zijn gevonden om dieper op in te gaan, is het onmogelijk om hier een volledige verklaring te geven. Daarom zal ik zoals aangekondigd de verklaringen van Peter Raedts voor de ontwikkeling van de negentiende-eeuwse middeleeuwen, bij met name de katholieken en protestanten, naast de grafieken leggen.

Raedts wijst in zijn boek een aantal publicaties en boeken aan die een impuls gaven voor interesse of juist negativiteit ten opzichte van de middeleeuwen. Als we deze boeken en publicaties op een tijdlijn plaatsen en die over de grafieken leggen, kunnen we zien of en hoe deze publicaties een verschil hebben gemaakt. Deze publicaties zijn opgesomd in tabel 3 hieronder, waarbij de letters onder de jaartallen verwijzen naar de letters in figuur 8 en 9. In de grafiek zijn kleine letters gebruikt om het einde van een bepaalde publicatiereeks te markeren.

⁵⁸ Idem, 245-254.

Tabel 3. Relevante publicaties over de middeleeuwen, zoals genoemd door Peter Raedts.

1846	Guillaume Groen van Prinsterer, <i>Handboek der geschiedenis van het vaderland</i>.
A	Orthodox protestants. In dit boek worden de middeleeuwen hoewel kort behandeld, gitzwart afgeschilderd. Dit boek was erg populair tot in de twintigste eeuw binnen protestantse kringen. ⁵⁹
1842-'50	<i>De Katholiek en De Godsdienstvriend, twee katholieke tijdschriften publiceren over de (katholieke) middeleeuwen.</i> Voornamelijk publicaties over kloosters en kerken vanuit de middeleeuwen tot heden, met opsommingen van namen en gebeurtenissen. ⁶⁰ Ook is er kritiek op protestantse pogingen om figuren als Thomas a Kempis voor te stellen als vroege protestanten, zoals hieronder.
1844-'53	Joseph Alberdingk Thijm, diverse gedichtenbundels met middeleeuwse gedichten, en historische (fictie) verhalen die zich in de middeleeuwen afspeelen. ⁶¹ Raedts vermeld enkel Thijm als invloedrijk persoon, maar geeft niet een datum of bepaald werk. ⁶² Thijm propageerde bovendien de neogotiek gedurende de tweede helft van de negentiende eeuw. ⁶³
1855	Petrus Hofstede de Groot, <i>De Groninger Godgeleerden in hunne eigenaardigheid</i>.
D	Al vanaf 1842 had de protestantse Hofstede de Groot ideeën over een bepaald uniek type de Nederlands christendom. Dit was niet per se katholiek, Luthers of Calvinistisch, maar een mengeling. Hierin paste ook de moderne devotie van Thomas a Kempis en Geert Grote. ⁶⁴ In deze publicatie, die gebaseerd was op een eerder toespraak, presenteerde hij zijn ideeën.
1864-'71	Willem Moll, <i>Kerkgeschiedenis van Nederland vóór de Hervorming</i>.
E	Eveneens een protestantse poging om de middeleeuwen niet als poel van ellende te presenteren, maar ook aandacht te geven aan de mooie kanten, ook van het christendom. Overigens was volgens hem de hervorming wel nodig voor echte vrijheid van het evangelie. ⁶⁵
1871-'72	Willem Jan Frans Nuyens, <i>Algemeen Geschiedenis des Nederlandschen Volks. Van de vroegste tijden tot op onze dagen</i>.
F	Waar de katholieken tot dan toe vooral kritiek hadden op de protestantse weergave van de middeleeuwen, of schreef over wat er van de middeleeuwen nog over was in het heden, deed de Katholiek Nuyens een poging om een werkelijke Nederlandse geschiedenis te schrijven waarin de middeleeuwen een prominente plek kregen. ⁶⁶ Dit boek zou de interesse in de middeleeuwen onder katholieken hebben doen opleven.

⁵⁹ Raedts, *De ontdekking*, 247-248.

⁶⁰ Idem, 255, 257.

⁶¹ DBNL, 'J.A. Alberdingk Thijm', <https://www.dbnl.org/auteurs/auteur.php?id=albe003> (geraadpleegd op 16 juni 2020).

⁶² Raedts, *De ontdekking*, 258-260.

⁶³ Ibidem.

⁶⁴ Idem, 244.

⁶⁵ Idem, 246.

⁶⁶ Idem, 256, 257.

Het is opvallend dat beide vroeg-katholieke publicaties (B,C) op de katholieken ook enige tijd vrijwel direct effect hebben gehad op de mate van negativiteit; deze daalde onmiddellijk. Overigens zwakt dit effect ook weer wat af na verloop van tijd. Bij de protestanten markeert het boek van Groen (A) eveneens een direct effect; hier is er een niet geheel onverwachtse stijging in negativiteit te zien. Maar, als de protestanten na minder dan tien jaar nog negatiever worden over de middeleeuwen (mogelijk groeide de populariteit van het boek in die periode, waardoor dit effect wat langer duurt), lijken de katholieken zich af te zetten tegen die protestantse negativiteit, en worden ze zelf minder negatief. Wellicht dat ook de liberalen door Groen's handboek geïnspireerd werden tot een meer negatievere houding, maar dit zou ook verklaard kunnen worden door de totstandkoming van de grondwet in deze periode. Liberalen moesten immers weinig hebben van middeleeuwse en feodale instelling zoals een grenzeloze koningschap.

Ook het boek van Moll (E) lijkt effect te hebben gehad; alleen niet in de groepen waar hij wellicht op hoopte, bij de protestanten of misschien bij de liberalen en neutralen, maar juist bij de katholieken. Hoewel zij zich niet konden vinden in de toch gematigde, en zelfs waarderende blik naar op de middeleeuwen waarbij prominente katholieken als voorlopers op de reformatie werden gezien, worden ze toch minder negatief. Mogelijk dat de kritiek op Moll en zijn middeleeuwen, gepaard gingen met interesse en onderzoek in hoe het dan wel zou zitten.

De publicatie van het boek van Nuyens (F) valt samen met het hoogtepunt van negativiteit. Dat hij daarmee ook verantwoordelijk is voor de algehele afname die op deze piek volgt, lijkt echter onwaarschijnlijk. Dat voor katholieken deze heeft bijgedragen aan een verminderde negativiteit, is uiteraard zeker wel mogelijk.

Overigens is het opmerkelijk genoeg niet zo dat het effect van de hiervoor genoemde de publicaties enkel wordt veroorzaakt doordat er veel over die publicaties werd geschreven. De namen van deze auteurs, of de titels van de boeken komen namelijk in de verkorte versies van de artikelen van 250 woorden die ik heb gebruikt, bijna niet voor in deze periodes; het blijft bij hooguit enkele keren per jaar. Het lijkt er daardoor op dat deze publicaties daadwerkelijk voor enige tijd na het publiceren ervan invloed hadden op het denken over de middeleeuwen.

De sterk verminderde negativiteit vanaf 1870-'75, is denk ik vooral het gevolg van enerzijds de toename van nieuwe genres, en anderzijds een verandering van denken over de

middeleeuwen, waarin men ook oog kreeg voor andere aspecten van deze periode, dan alleen het katholicisme en het kennisverschil met de mens van toen. In de krantenartikelen vanaf 1875 zag ik een toename van interesse in de kunst en cultuur van de middeleeuwen, waarbij de middeleeuwen werden gezien als een meer 'normale' historische periode. Dit was niet alleen in de academie, maar ook daarbuiten. Er waren bijvoorbeeld publieke lezingen en er verschenen boeken over onder andere middeleeuwse filosofie, recht, letterkunde en kunst. Door de toename van nieuwe genres was er meer ruimte om dit soort boeken en lezingen, of zelfs wetenschappelijke onderzoeken te bespreken, maar ook voor informatieve historische achtergronden, of reisverhalen waarin op romantische wijze wordt gesproken over middeleeuwse steden en dorpjes.

8. Conclusie

De verklaringen voor de ontwikkeling van de houding tot de middeleeuwen, ligt dus deels in een aantal publicatie van zowel protestanten als katholieken, waarbij deze publicaties ook reacties lijken uit te lokken bij elkaar. Toch is nog dieper nader onderzoek nodig, naar bijvoorbeeld de vraag hoe het kan dat na 1875 de negativiteit zo af lijkt te nemen. Hiervoor zou een model die de artikelen ook op onderwerp sorteert erg behulpzaam zijn, om er zo achter te komen naar waar er gezocht moet worden. Ook zou er nog onderzoek gedaan kunnen worden naar of en hoe de academische interesse in de middeleeuwen, waarover geschreven wordt in de kranten, invloed had op de houding tegenover de middeleeuwen. In ieder geval kunnen we hier wel zeker concluderen dat het in het laatste kwart van de negentiende eeuw de middeleeuwen toch in zekere mate werden gewaardeerd.

Conclusie

In dit onderzoek heb ik laten zien hoe historici gebruik kunnen maken van *neural network*-algoritmes bij de analyse van grote hoeveelheden gedigitaliseerde bronnen, in dit geval krantenartikelen, waarbij deze gesorteerd werden op de houding die ze tegenover de middeleeuwen hadden. Ik heb betoogd dat er geen fundamentele theoretische bezwaren zijn om deze technologie te gebruiken, waarbij ik ook heb uitgelegd waarom deze techniek beter is dan andere geautomatiseerde methodes of algoritmes voor de analyse van historische bronnen. Vanwege de toegankelijkheid van *deep learning*-algoritmes, zowel wat betreft mogelijkheden voor zelf-educatie, als de daadwerkelijk benodigde technologie, is het ook voor historici mogelijk hiervan gebruik te maken.

Het is overigens belangrijk om te benadrukken dat deze manier van analyse niet bedoelt is als vervanging van een enige andere analoge methode, maar vooral een aanvulling. Voor het trainen van een algoritme is het bovendien noodzakelijk dat de historicus eerst een groot aantal van de bronnen zelf beoordeeld, waarna het model in feite het werk van de historicus verder uitbreid over de rest van de bronnen. Bovendien is een groot voordeel hiervan dat de data bepaalde ontwikkelingen zichtbaar kan maken die anders mogelijk verborgen zouden blijven, en daarmee als startpunt dienen voor verder kwalitatief onderzoek.

In het praktische gedeelte van dit onderzoek heb ik dat laten zien, waarbij ik onderzocht hoe tussen 1830-1910 de houding binnen de Nederlandse samenleving ten opzichte van de middeleeuwen ontwikkelde, afgeleid uit het gebruik van de middeleeuwen krantenartikelen in die periode. Hierin werd duidelijk dat er in de ontwikkeling van de houding tot de

middeleeuwen gedurende de negentiende eeuw een aantal opmerkelijke schommelingen waren. Er waren drie pieken van negativiteit rond 1840, 1850, en 1870. In die pieken waren er echter wel grote verschillen tussen voornamelijk de katholieken aan de ene kant, en de neutralen en liberalen aan de andere kant.

Deze verschillen, evenals de schommelingen, leken veroorzaakt te worden door een aantal publicaties die Peter Raedts als belangrijk had aangemerkt. Bovendien wekten de resultaten de indruk dat vooral de katholieken en liberalen op elkaar reageerden; daar waar liberalen negatiever werden, werden de katholieken juist minder negatief. Daarnaast blijkt er een opvallende afname van negativiteit te zien voor alle zuilen vanaf de periode 1870-'75. Deze afname is mogelijk het gevolg van onder andere een toenemende academische interesse in de middeleeuwen, die ook de kranten bereikte, waarin de middeleeuwen niet meer zozeer gebruikt werden als katholieke stroman voor de protestanten en liberalen, maar er ook aandacht was voor bijvoorbeeld de letterkunde uit deze periode.

Verder onderzoek zou zich kunnen richten deze periode vanaf 1870, evenals op de wisselwerking tussen de katholieke en liberale kranten in de periode daarvoor. Daarnaast zou het interessant zijn naar verschillen en overeenkomsten tussen regio's, maar ook tussen conservatieve, progressieve en eventueel meer sociale vleugels van de zuilen.

Wat sommigen wellicht zien of zagen als een duel tussen de kwalitatieve, geautomatiseerde, en de kwalitatieve hermeneutisch methode, is dat dus eigenlijk niet. Als het dan toch een duel is, hoop ik op een verloop en uitkomst zoals van het duel tussen de baron Von Denkhäus en de heer Von Bounhart: *beiden streden met de grootste hardnekkigheid*, waarbij de heren, na elkaar enkele kwetsuren te hebben toegebracht en *de heelmeeesters beiden met hunne hulp hadden bijgestaan, de voorzitter des hofs beide partijen uitnoodigde, zich te verzoenen, hetgeen zij aanstonds deden, door elkander de hand te geven en te omhelzen.*⁶⁷

⁶⁷ 'Deutschland', *Leeuwarder Courant*.

Bibliografie

- ‘Duitschland’, *Leeuwarder courant* (Leeuwarden 14 juli 1846) 1 (geraadpleegd op 3-7-2020 via Delpher, <http://resolver.kb.nl/resolve?urn=ddd:010583417:mpeg21:p001>)
- Apostolova, Emilia, en R Andrew Kreek, ‘Training and Prediction Data Discrepancies: Challenges of Text Classification with Noisy, Historical Data’, in: *Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text* (Brussel 2018) 104–109 <doi:10.18653/v1/w18-6114>.
- Benjamini, Yoav, en Yosef Hochberg, ‘Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing’, *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1995) 289–300.
- Bergstra, J, D Yamins, en D D Cox, ‘Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures’, in: *Proceedings of the 30th International Conference on Machine Learning* 28 (Atlanta, Georgia 2013).
- Blom, J. C. H., en J.J. Cahen, ‘Jewish Netherlanders. Netherlands Jews, and Jews in the Netherlands, 1870–1940’, in: J. C. H.; Blom, R. G.; Fuks-Mansfeld en I. Schoffer ed., *History of the Jews in the Netherlands* (Portland, Oregon 2001) 230–295.
- Blondé, B e.a., *Trend en toeval. Inleiding tot de kwantitatieve methoden voor historici*. (Leuven 2015).

- Boldt, Andreas, 'Perception, Depiction and Description of European History: Leopold von Ranke and his Development and Understanding of Modern Historical Writing', *eSharp* (2007) 1–17.
- Broersma, Marcel, en Frank Harbers, 'Exploring Machine Learning to Study the Long-Term Transformation of News: Digital newspaper archives, journalism history, and algorithmic transparency', *Digital Journalism* 6 (2018) 1150–1164 <doi:10.1080/21670811.2018.1513337>.
- DBNL, 'J.A. Alberdingk Thijm' <https://www.dbnl.org/auteurs/auteur.php?id=albe003> (geraadpleegd op 2020-06-17).
- Delpher, 'Wat zit er in Delpher?', <https://www.delpher.nl/nl/platform/pages/helpitems?nid=395> (geraadpleegd op 2020-07-06.)
- Eijnatten, Joris van, Toine Pieters, en Jaap Verheul, 'Big Data for Global History. The Transformative Promise of Digital Humanities', *BMGN - Low Countries Historical Review* 128 (2013) 55–77.
- De Graaf, R.H., 'Journalistiek in beweging. Veranderende berichtgeving in kranten en pamfletten (Groningen en 's-Hertogenbosch 1813-1899)' (Proefschrift Utrecht 2009).
- Graham, Shawn, Ian Milligan, en Scott Weingart, *Exploring big historical data : the historian's macroscope* (Londen 2016).
- Hockey, Susan, 'The History of Humanities Computing', in: Susan Schreibman, Ray Siemens en John Unsworth ed., *A Companion to Digital Humanities* (1ste druk; Oxford 2004) 1–19 <doi:10.1002/9780470999875.ch1>.
- Jockers, Matthew L., en Ted Underwood, 'Text-Mining the Humanities', in: Susan Schreibman, Ray Siemens en John Unsworth ed., *A new Companion to Digital Humanities* (Oxford 2016) 291–306.
- Kettunen, Kimmo, 'Ground Truth OCR Sample Data of Finnish Historical Newspapers and Journals in Data Improvement Validation of a re-OCRing Process', *Liber Quarterly* 30 1–20 <doi:10.18352/lq.10322>.
- Meijer, Jaap, 'Arduous adaptation, 1814-1870', in: J. C. H.; Blom, R. G.; Fuks-Mansfeld en I. Schoffer ed., *History of the Jews in the Netherlands* (Portland, Oregon 2001) 192–229.
- Neuendorf, Kimberly A., *The content analysis guidebook* (2de druk; Los Angeles 2017).
- Noymer, Andrew, 'Alpha, Significance Level of Test', *Encyclopedia of Survey Research Methods* (Thousand Oaks 2011) 17–19.

Strange, Carolyn e.a., 'Mining for the Meanings of a Murder: The Impact of OCR Quality on the Use of Digitized Historical Newspapers', *Digital Humanities Quarterly* 8 (2014).

Swierenga, Robert P., 'Clio and computers: A survey of computerized research in history', *Computers and the Humanities* 5 (1970) 1–21 <doi:10.1007/BF02404252>.

Tulkens, Stéphan, Chris Emmery, en Walter Daelemans, 'Evaluating unsupervised Dutch Word embeddings as a linguistic resource', *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016* (2016) 4130–4136.

Wielenga, Friso, *De geschiedenis van Nederland. Van de Opstand tot heden* (Amsterdam 2012).

Wijfjes, Huub, 'Digital Humanities and Media History. A Challenge for Historical Newspaper Research', *Tijdschrift voor Mediageschiedenis* 20 (2017) 4–24.

Zwaan, Janneke van der, en Lotte Wilms, 'Ochre. A Toolbox for OCR Post-Correction' (Koninklijke Bibliotheek, 2018) <doi:http://doi.org/10.5281/zenodo.1189245>.

Appendices

A. Codelijst

Voordat ik aan het onderzoek kan beginnen, moet ik vaststellen wat ik ga beoordelen als een positieve, neutrale en negatieve houding ten opzichte van de Middeleeuwen. Net zo belangrijk is het antwoord op de vraag hoe er met de citaten of parafrases in artikelen omgegaan moet worden. Is een artikel waarin iemand geciteerd wordt die spreekt over 'de duistere Middeleeuwen', maar dat dit gebruik van 'Middeleeuwen' door het artikel wordt veroordeeld, positief of negatief, of neutraal? En een lezing over de Middeleeuwen waarover positief wordt gesproken, duidt dat ook op een positieve houding? Of ligt dat juist weer aan de inhoud van de lezing of andere redenen waarom de schrijver van het artikel zegt positief te zijn? Hieronder zal ik in deze dingen meer duidelijkheid scheppen.

Positief

Als positief beoordeeld, zijn artikelen waarin het woord 'Middeleeuwen' of 'middeleeuws(e)' direct, ofwel indirect gebruikt wordt als:

- Een ideaal of iets wat (erg) mooi is. Met name bij het gebruik als bijvoeglijk naamwoord, zoals een 'mooi middeleeuws gebouw';
- Een betere tijd dan de huidige, met grotendeels weglating van nuance, zoals 'de arbeiders in de middeleeuwse gilden waren beter af dan de arbeiders in deze tijd'.
- Een algehele positieve houding tegenover dat wat uit de Middeleeuwen komt, of dat wat als middeleeuws wordt gezien: een meer romantische kijk.
- In sommige gevallen een aankondiging voor een boek of lezing over de middeleeuwen, waarbij duidelijk is dat het boek of of de lezing een positieve kijk op de middeleeuwen geeft. Of als er een positieve houding ten opzichte van de middeleeuwen in de aankondiging zit. Het idee hierachter is dat iemand een boek aanraad als hij denkt dat degene waaraan het wordt aangeraden, dat boek mogelijk interessant vind, of het enigszins met de inhoud eens zou zijn.
- Duidelijke interesse in de middeleeuwen, met oog voor wat goed of mooi was in de middeleeuwen.

Neutraal

Als neutraal beoordeeld, zijn artikelen waarin het woord 'Middeleeuwen' of 'middeleeuws(e)' direct, ofwel indirect gebruikt wordt op deze manier:

- Er wordt een realistische, genuanceerd beeld geschetst. Er waren positieve en negatieve elementen;
- Een neutrale constatering dat iets uit de Middeleeuwen stamt of middeleeuws is, zonder daarmee iets positiefs of negatiefs te willen impliceren, bijvoorbeeld de melding dat een land sinds de Middeleeuwen een koning heeft.
- Een aankondiging voor een boek of lezing over de middeleeuwen waarbij het boek een neutrale houding ten opzicht van de middeleeuwen lijkt te hebben, of als uit de aankondiging blijkt dat de schrijver ervan een meer neutrale houding lijkt te hebben ten opzichte van de middeleeuwen.
- Een universitaire of anderszins publieke aanstelling, of een soortgelijke aankondiging, waarbij het onderwerp de middeleeuwen is, zonder daaraan een waarde oordeel aan te geven (betrekking hebbend op de Middeleeuwen)

Negatief

Als negatief beoordeeld, zijn artikelen waarin het woord 'Middeleeuwen' of 'middeleeuws(e)' direct, ofwel indirect gebruikt wordt als:

- Iets wat slecht, verwerpelijk of dom is. Met name bij het gebruik als bijvoeglijk naamwoord, zoals 'dit zijn middeleeuwse toestanden';
- Een slechtere tijd dan de huidige, met grotendeels weglating van nuance, zoals 'in de middeleeuwen dacht men niet na, maar in deze tijd is dat niet meer zo';
- Een algehele negatieve houding tegenover dat wat uit de Middeleeuwen komt, of dat wat als middeleeuws wordt gezien: een meer verlichte blik ten aanzien van de duistere middeleeuwen.
- In sommige gevallen een aankondiging voor een boek of lezing over de middeleeuwen, waarbij duidelijk is dat het boek of of de lezing een negatieve kijk op de middeleeuwen geeft. Of als er een negatieve houding ten opzichte van de middeleeuwen in de aankondiging zit.

Citaten en parafrases

De eigenlijke vraag is, of ik mij richt op wat de krant vindt, of wat de mensen in de tijd vonden. Daar ik mij met name op de kranten en hun lezers richt, en de krant gebruik als representatie van een bepaalde bevolkingsgroep, zal de mening van de krant leidend zijn. Dat ziet er als volgt uit:

Hoewel eerder genoemde algemene criteria voor het beoordelen van artikelen in veel gevallen zullen volstaan, is er toch enige aanvulling hierop noodzakelijk.

Positief

- Een parafrase of citaat die positief over de middeleeuwen spreekt, waarbij de schrijver met het beweerde, ofwel met woorden, ofwel juist stilzwijgend, instemt;
- Een parafrase of citaat die negatief over de middeleeuwen spreekt, waarbij de schrijver het beweerde zodanig twijfel trekt of bestrijd, dat het een positieve houding ten opzichte van de middeleeuwen verraad.

Neutraal

- Een parafrase of citaat die neutraal of genuanceerd over de middeleeuwen spreekt, waarbij de schrijver met het beweerde, ofwel met woorden, ofwel juist stilzwijgend instemt;
- Een parafrase of citaat die positief of negatief over de middeleeuwen spreekt, waarbij de schrijver het beweerde in twijfel trekt, of althans enigszins afzwakt.

Negatief

- Een parafrase of citaat die negatief over de middeleeuwen spreekt, waarbij de schrijver met het beweerde, ofwel met woorden, ofwel juist stilzwijgend, instemt;
- Een parafrase of citaat die positief over de middeleeuwen spreekt, waarbij de schrijver het beweerde zodanig twijfel trekt of bestrijd, dat het een negatieve houding ten opzichte van de middeleeuwen verraad.

B. Model

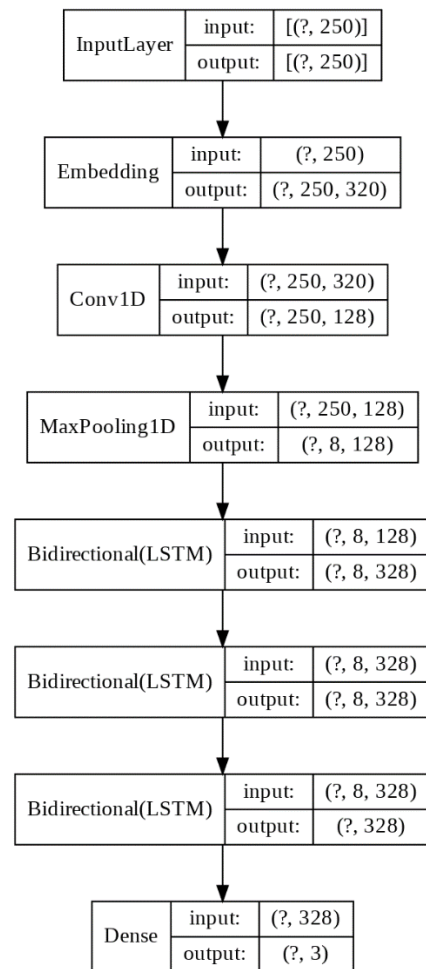
Tabel 4. De belangrijkste instellingen van het model.

Conv1D filters	128
Conv1D kernel size	7
MaxPooling1D size	28
MaxPooling1D stride	28
BiLSTM lagen	3
BiLSTM cellen	164
BiLSTM i1	0.2
BiLSTM i2	0.5
BiLSTM recurrent dropout	0.4
Output activation	Softmax
Optimizer	Adam
Label smoothing	0.2
Batch size	16
Epochs	8

Opmerkingen

De belangrijkste instellingen, zoals het aantal lagen en de grootte daarvan, van dit model heb ik gevonden met het Hyperopt-zoekalgoritme.⁶⁸ Zelf heb ik vervolgens geprobeerd met het aanpassen van bepaalde instellingen een beter model te krijgen, waarvan sommige aanpassingen ook dit model hebben gehaald.

Ik heb het model getraind met een gebalanceerde dataset, waardoor per batch het aantal voorbeelden per categorie ongeveer gelijk was, om zo een *bias* voor de neutrale categorie waarvan er veel meer voorbeelden waren, zoveel mogelijk te beperken.



Figuur 13. Het model met de *input* en *output shapes* per laag.

⁶⁸ J Bergstra, D Yamins en D D Cox, 'Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures', in: *Proceedings of the 30th International Conference on Machine Learning* 28 (Atlanta, Georgia 2013).