

Een aanbevelingssysteem voor juridische documenten

Arnoud De Jonge - 01808870

Arno Vermote - 01806792

Inhoud

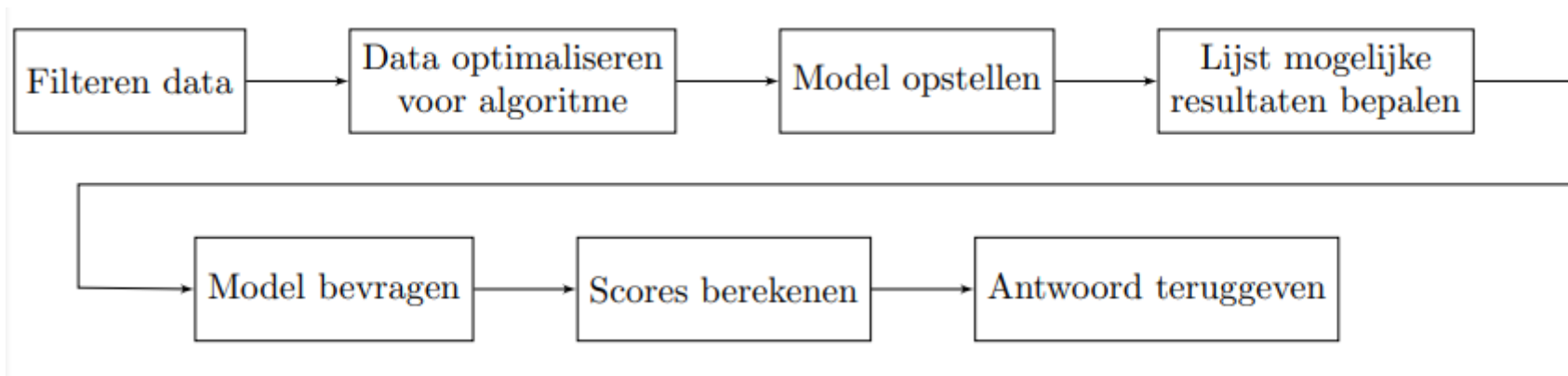
- Onze ideeën
- Eindconclusie
- Demonstratie

Word2Vec, Doc2Vec en BERT

- Pijplijn
- Gebruikte wetteksten
- Gewogen score

	(hoger is beter) Silhouette score	(hoger is beter) Dunn Index	(dichter bij 0 is beter) S_Dbw	(dichter bij 0 is beter) Davies-Bouldin score
TF-IDF	-0.16536	2.59E-05	0.82416	4.6662
Doc2Vec	-0.14014	1.5727E-05	0.367126	3.2904
Word2Vec	-0.077212	2.96568E-05	0.34176	1.7568
LDA	-0.14586	0.000003168	0.426	2.11464
Dispersion	-0.21879	4.62073E-05	1.6544	57.985704

Tabel 1: Clusteranalyse van verschillende NLP-technieken
Groen is beste, geel is 2de beste



Eindconclusie

- Werkt (zie demonstratie)
 - BERT sprookjes vergelijken
 - Word2Vec & Doc2Vec uitvoeren
 - Resultaten bespreken
- Wel traag (om lokaal te gebruiken)
 - BERT met juridische teksten gaat niet live