

Workshop Text - Mining

Berend Broesder & Sami Mohamed
Laurens Wijnholds & Maarten Ploemacher



Waar denken jullie aan bij text-mining?

Introductie

Waar denk je aan
bij text-mining?

Wat is text-mining?

Hoe werkt text-
mining?

Toekomstbeeld van
text-mining

Voor- & nadelen

Discussie



Wat is text-mining nou echt?

“Het omzetten van ongestructureerde teksten naar gestructureerde inzichten met behulp van automatisering.”



Welke soorten van text-mining bestaan er?

- **Clustering**

- Groepering

- **Classification**

- Labelen

- **Sentimentanalyse**

- Emotionele toon

- **Topic modeling**

- Onderwerp & belangrijke woorden uit de tekst halen

- **Information clustering (IE)**

- Herkenning van type tekst (locaties, namen, datums)

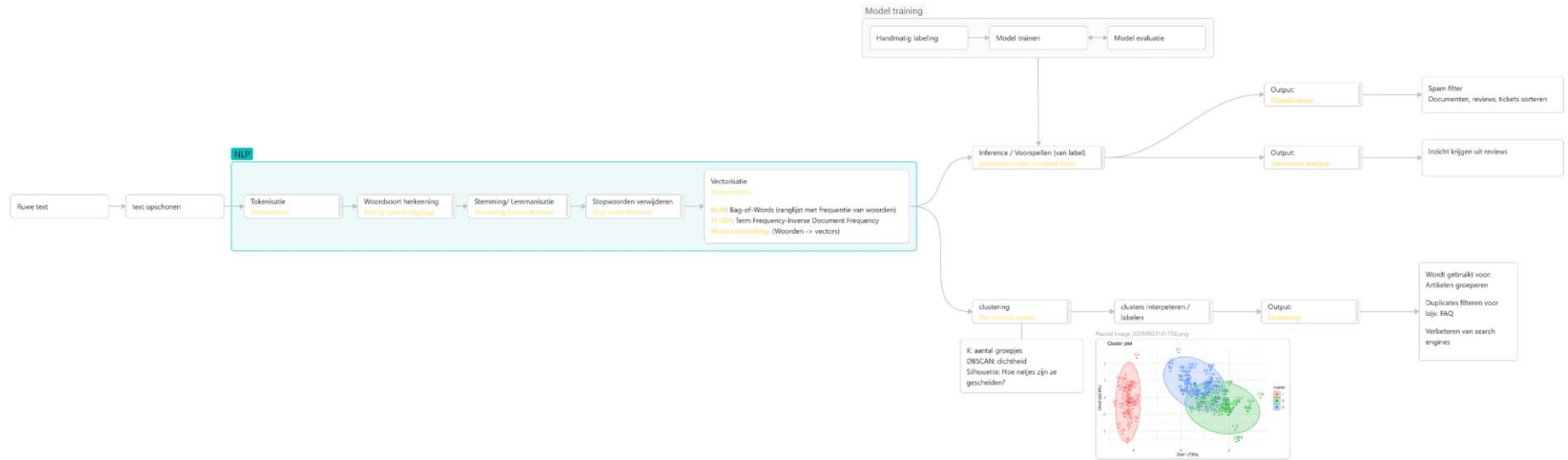
- **Information retrieval (IR)**

- Zoeken van documenten bij een zoekopdracht

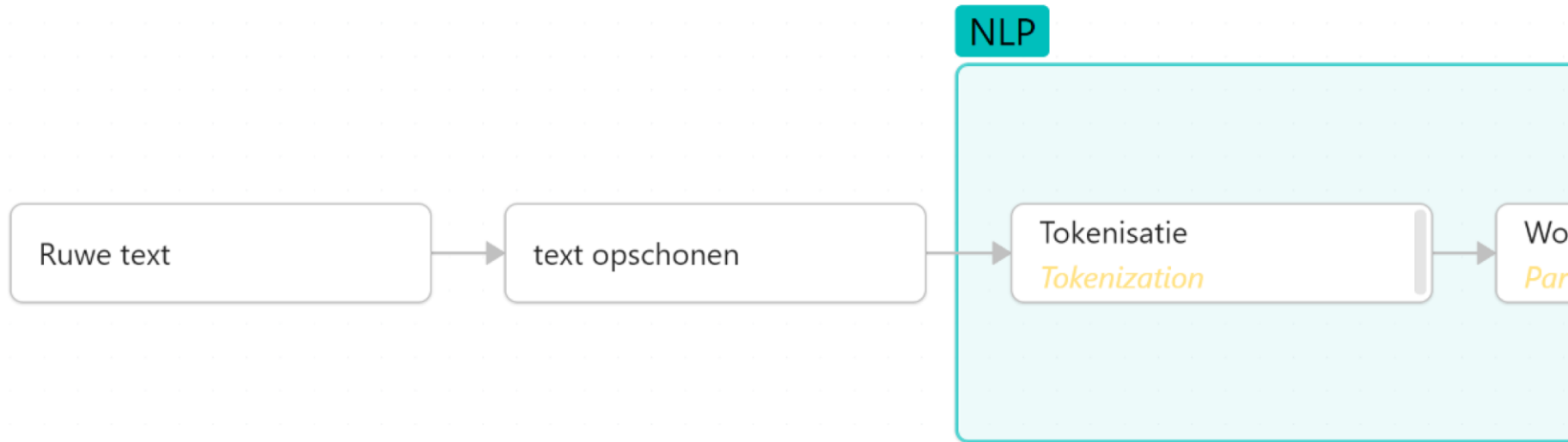
- **Association**

- Verbanden vinden tussen woorden

Hoe werkt text-mining dan?



Hoe werkt text-mining dan?



Hoe werkt text-mining dan?

NLP

Tokenisatie
Tokenization

Woordsoort herkenning
Part of speech tagging

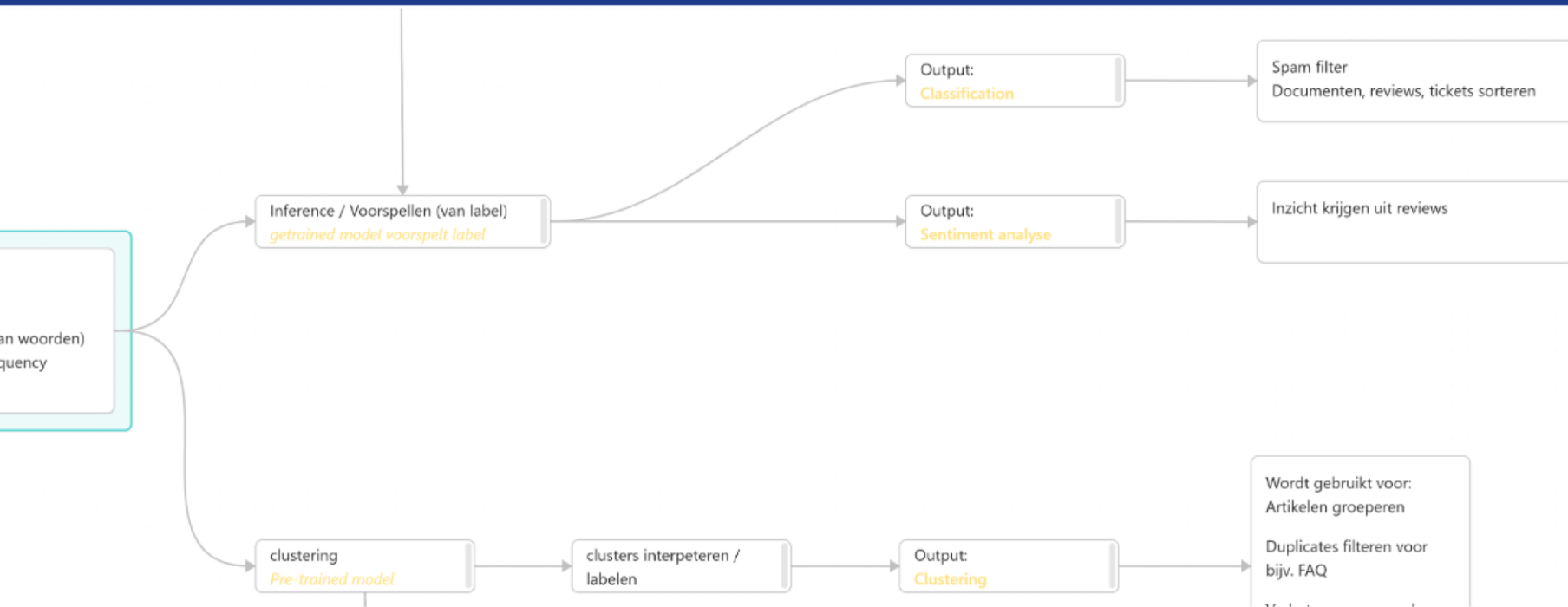
Stemming/ Lemmanisatie
Stemming/Lemmatization

Stopwoorden verwijderen
Stop-word Removal

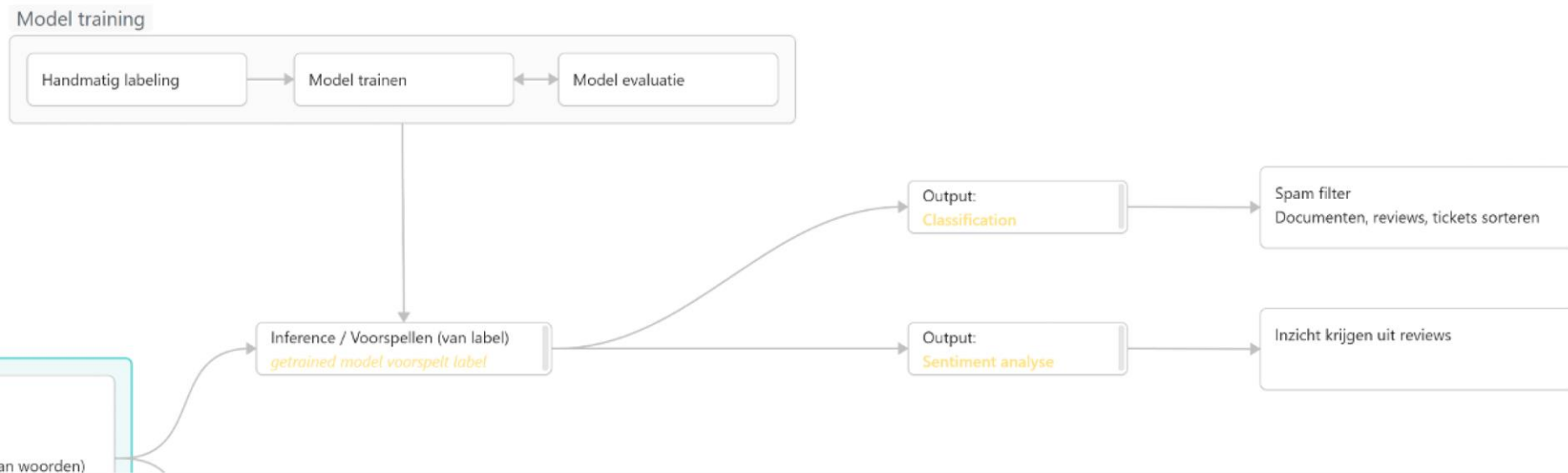
Vectorisatie
Vectorization

BoW: Bag-of-Words (ranglijst met frequentie van woorden)
TF-IDF: Term Frequency-Inverse Document Frequency
Word Embeddings (Woorden -> vectors)

Hoe werkt text-mining dan?



Hoe werkt text-mining dan?



Hoe werkt text-mining dan?

clustering
Pre-trained model

K: aantal groepjes
DBSCAN: dichtheid
Silhouette: Hoe netjes zijn ze
gescheiden?

clusters interpreteren /
labelen

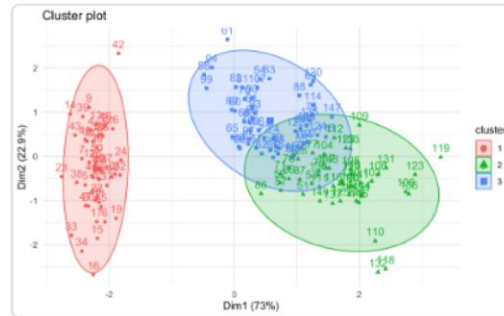
Output:
Clustering

Wordt gebruikt voor:
Artikelen groeperen

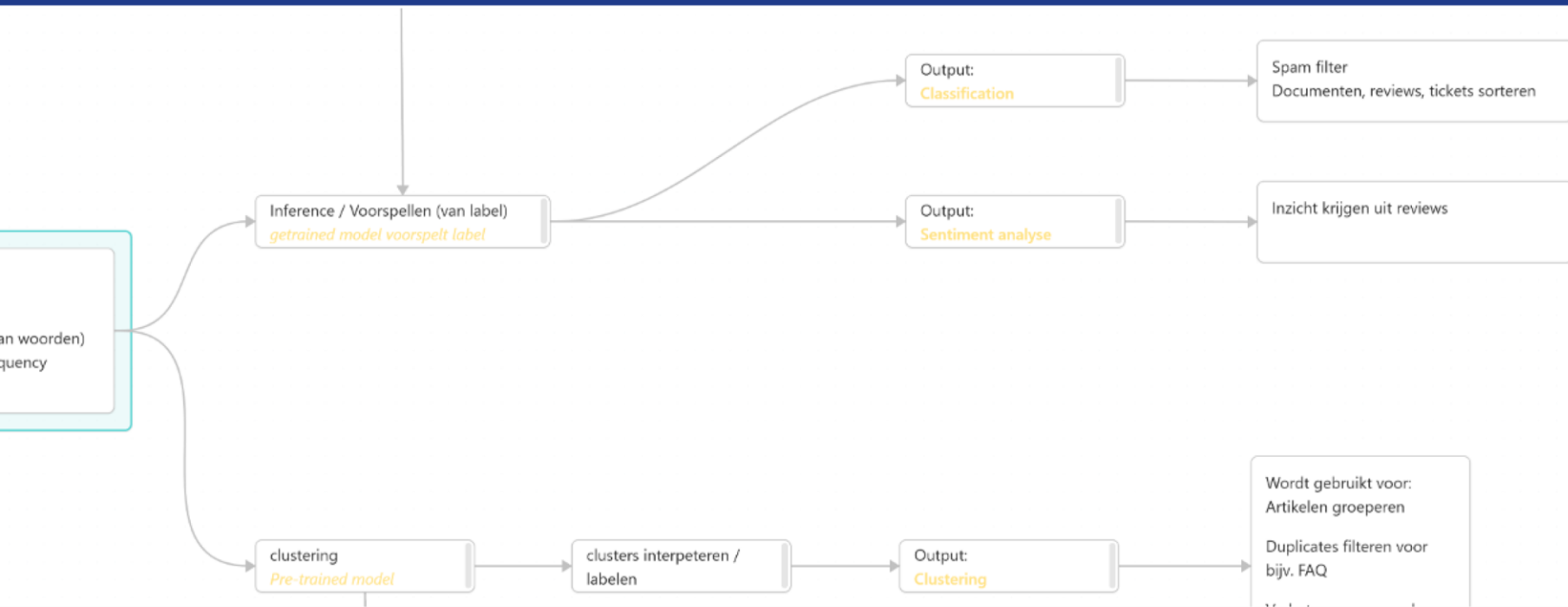
Duplicates filteren voor
bijv. FAQ

Verbeteren van search
engines

Pasted image 20250923101758.png



Hoe werkt text-mining dan?





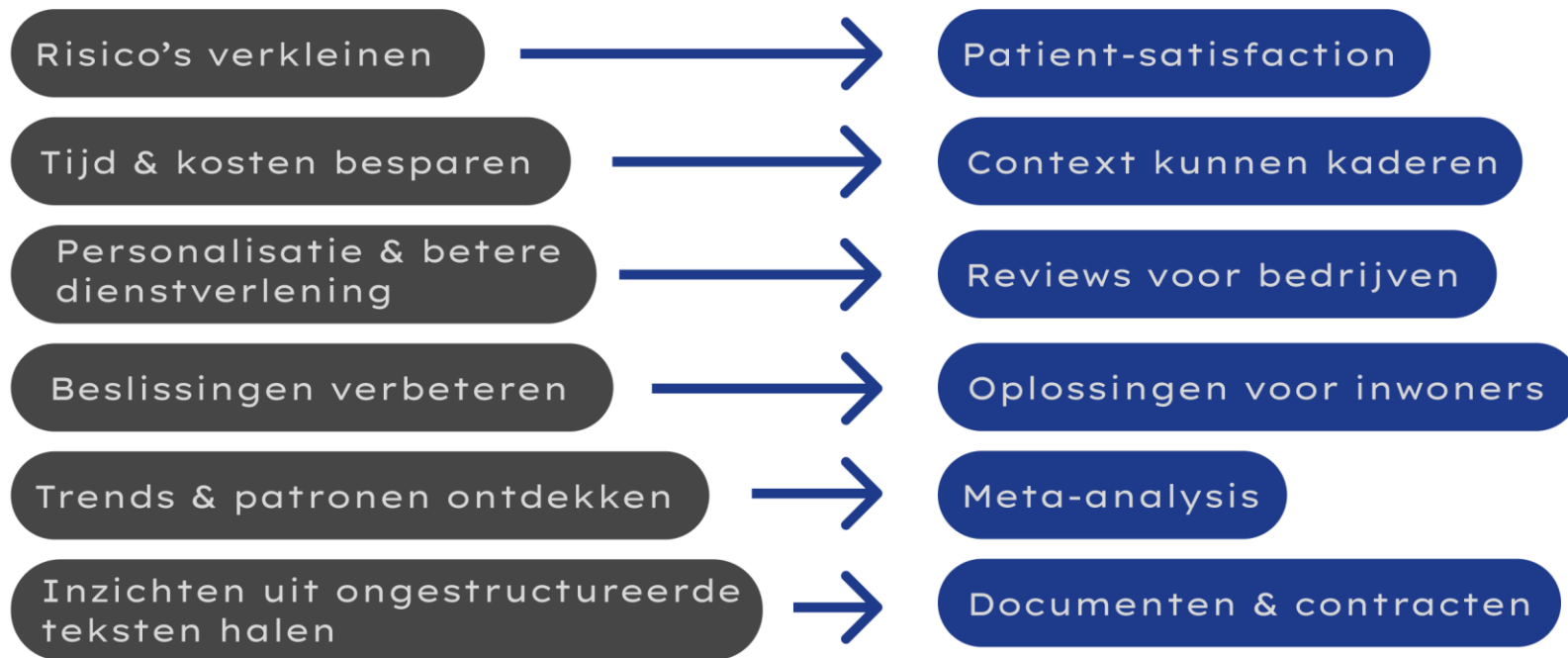
Toekomstbeeld van text-mining

Geautomatiseerde controle van berichten

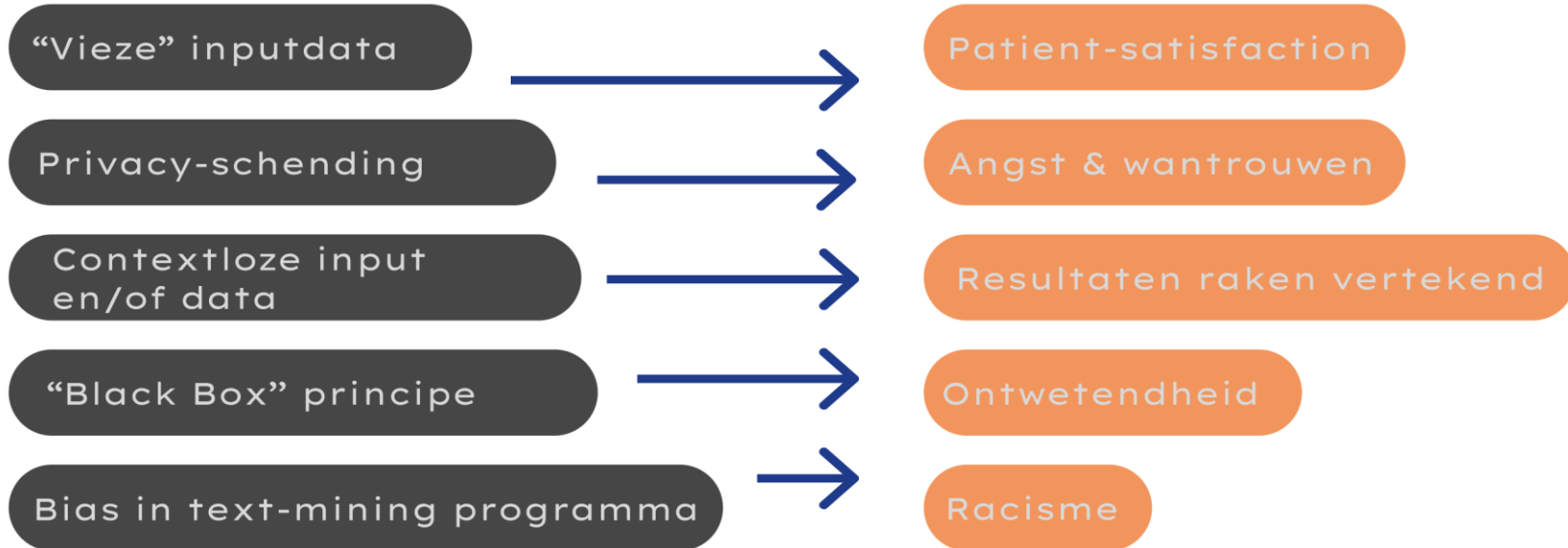
Hoe denkt de wereld over jou specifiek?

LLM's vs Text-mining: Zwitsers zakmes vs scalpel

Voordelen & voorbeelden



Nadelen & voorbeelden





Discussie

Helemaal niet
mee eens

Beetje niet mee
eens

Beetje wel mee
eens

Helemaal wel
mee eens

"Kostenbesparing rechtvaardigt de inzet van
AI voor text-mining protocollen voor
contentmoderatie."



Discussie

Helemaal niet
mee eens

Beetje niet mee
eens

Beetje wel mee
eens

Helemaal wel
mee eens

"Mogen bedrijven de (gevoelige) inhoud van emails gebruiken om het spam filter te verbeteren?"

Afronding



Neem onze handout mee!

Hoe helpt dit je Datastory?

Wat vond jij het interessantst?

Wat denk je nu dat text-mining is?