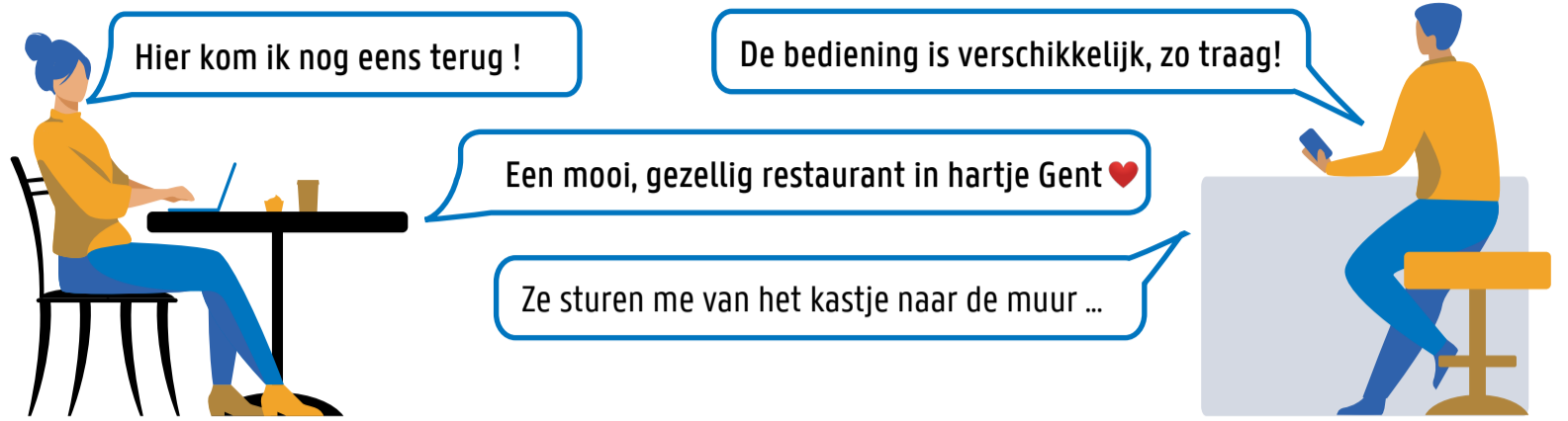


AUTOMATISCHE SENTIMENTANALYSE IN TEKSTEN

Input

De invoertekst voor automatische sentimentanalyse: klantenreviews.



Preprocessing

Inputtekst wordt voorbereid voor automatische verwerking.

Preprocessing kan op dezelfde manier gebeuren voor lexicon-gebaseerde en machinelearning-aanpakken.

Lowercasen

Alle tekst omzetten naar kleine letters ("Gezellig resto in Gent" → "gezellig resto in gent")

Tokeniseren

Zinnen in woorden splitsen ("gezellig resto in gent" → "gezellig", "resto", "in", "gent")

Lemmatiseren

Woorden vervangen door lemma/woordenboekvorm ("is" --> "zijn", "hartje" → "hart")

Part-of-speech tagging

Automatisch woordsoorten toekennen ("zijn" = werkwoord, "hart" = zelfstandig naamwoord)

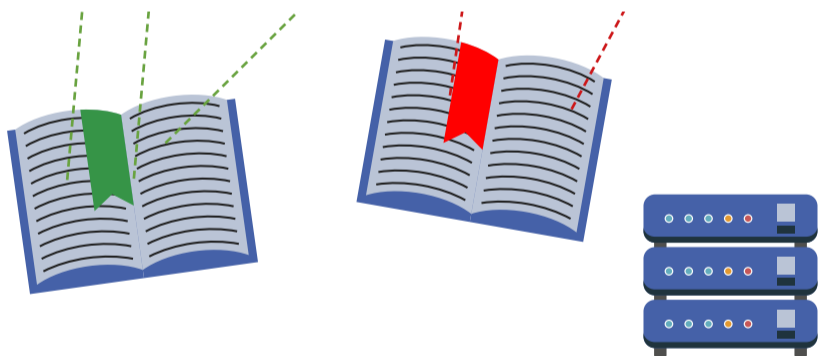
Lexicongebaseerd (woordenlijsten)

Machinelearning (lerende aanpak)

Woordenlijsten + matching

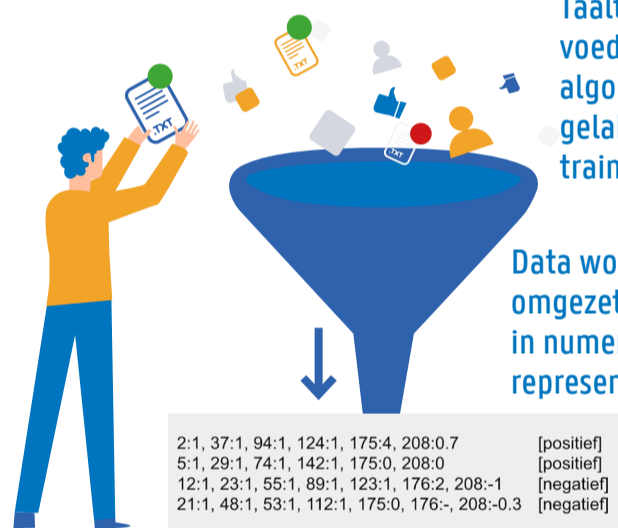
Een **mooi, gezellig** restaurant in hartje Gent ❤️
De bediening is **verschikkelijk, zo traag!**

POSITIEVE, NEGATIEVE en **NEUTRALE** woorden worden gematched met woordenlijsten en een globale sentimentscore wordt berekend voor de tekst.



Taaltechnologen voeden algoritmes met gelabelde trainingsdata.

Data worden omgezet in numerieke representaties.



Trainen van een model met features

Classificatie

Een **mooi, gezellig** restaurant in hartje Gent ❤️ **+3**

De bediening is **verschikkelijk, zo traag!** **-2**

2:1, 3:7, 74:1, 124:5, 205:0,7 [positief|negatief|neutraal] ?

Nieuwe, te voorspellen teksten worden omgezet in numerieke representaties zodat het getrainde model een voorspelling kan doen.

Heerlijke pasta in een gezellige setting ❤️

Output

Lexicongebaseerde Sentimentvoorspelling

POSITIEF
NEUTRAAL
POSITIEF

TE VOORSPELLEN TEKST

Heerlijke pasta in een gezellige setting ❤️
15 EUR voor een Dame Blanche?!
De stroom **viel** even uit, maar dit werd **goed** opgevangen door het personeel.

Machinelearning Sentimentvoorspelling

POSITIEF
NEGATIEF
POSITIEF

WWW.LT3.UGENT.BE/SENTIMENT-DEMO