



university of
 groningen

faculty of economics
 and business

customer insights center

Nieuwe inzichten van geschreven Big Data

Tekstmining voor Marketing

Dr. Evert de Haan

Report RUGCIC-2020-02
ISBN 978-94-034-2900-7



Inhoudsopgave

✓ Management samenvatting	Pagina 3
✓ Waarom tekst mining?	Pagina 13
✓ Bronnen van geschreven data	Pagina 18
✓ Inzichten verkregen van geschreven data	Pagina 21
✓ Sentimentanalyse	Pagina 23
✓ Onderwerp analyse	Pagina 35
✓ De relatie met de prestatie	Pagina 45
✓ Integratie in een klantendashboard	Pagina 59
✓ Curriculum Vitae van de auteur	Pagina 65
✓ Referenties	Pagina 67
✓ Customer Insights Center	Pagina 70



university of
 groningen

faculty of economics
 and business

customer insights center

Management samenvatting



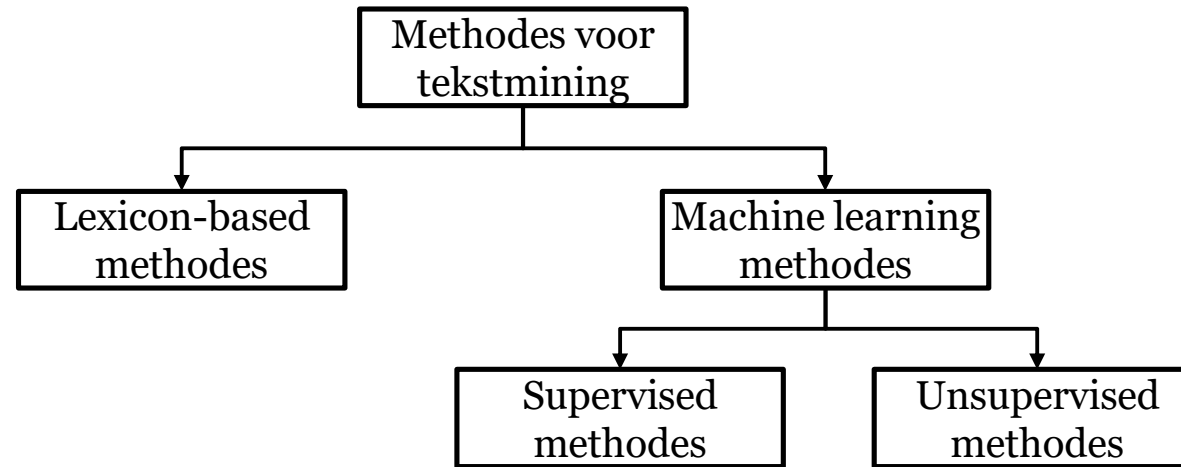


Wat is tekstmining?

- › Tekstmining kan gedefinieerd worden als “het achterhalen van nieuwe, voorheen onbekende, informatie uit geschreven bronnen door een geautomatiseerd computerproces”¹
- › In plaats van handmatig door geschreven stukken van informatie te gaan en deze classificeren, b.v. klantenreviews en sociale media discussies, kun je dit door een algoritme laten doen.
- › Voordelen zijn dat algoritmes objectief zijn (bv. een stukje tekst wordt iedere keer hetzelfde geclassificeerd), ze kunnen meer informatie aan in een fractie van de tijd dat mensen dit kunnen, en het is veel goedkoper.
- › Nadelen zijn dat algoritmes een tekst niet écht kunnen begrijpen; het weet de context niet en begrijpt sarcasme en cynisme niet.
- › In dit rapport worden tekstmining methoden, kansen en uitdagingen in de marketing besproken.
- › ¹ <http://people.ischool.berkeley.edu/~hearst/text-mining.html>



Voorbeelden aanpak tekst mining



Taak

Tekst classificatie

- > Sentiment
- > Content (met unieke klassen)

Onderwerp modellering

Voorbeeld methodes

✓	✓	X
X	✓	X
X	X	✓
LIWC	SVM	LDA



Wat betekent ...?

- › **Latent Dirichlet Allocation (LDA):** Een machine learning techniek gebruikt om in een stuk tekst de onderliggende onderwerpen te ontdekken en te classificeren.
- › **Linguistic Inquiry en Word Count (LIWC):** Een populair programma voor tekstanalyse, gebruikt voor het tellen van woorden en (basis) sentiment analyse.
- › **Machine learning (ML):** Een subset van Artificial Intelligence. ML methodes gebruiken een training dataset om patronen te ontdekken, welke vervolgens gebruikt kunnen worden voor (automatische) voorspellingen. Dit kan zijn:
 - **Supervised ML:** ML algoritmes krijgen trainingsdata (bv. stukjes tekst) samen met geschatte en gewenste output (bv. hand gecodeerde sentimenten)
 - **Unsupervised ML:** ML algoritmes herkennen patronen zelf zonder voor-gedefinieerde output, bv. een ML algoritme wat zichzelf leert dat soortgelijke woorden bij hetzelfde onderwerp horen.



Wat betekent ...?

- › **Natural Language Processing (NLP):** Een gebied wat gaat over de interactie tussen computers en mensen (natuurlijke) talen.
- › **Polarity (polariteit):** De hoeveelheid van een mening in een document of stukje tekst wat positief, negatief, gemixt of neutraal is.
- › **Support Vector Machines (SVM):** Een (supervised) machine learning methode welke gebruikt kan worden voor het classificeren van data (bv. voor sentiment of content).
- › **Sentiment analyse:** Een set van NLP technieken welke de polariteit of andere emoties van een (geschreven) stuk informatie proberen te vinden.
- › **Onderwerp analyse:** Een aanpak om onderliggende onderwerpen van een collectie van teksten of documenten te ontdekken en te classificeren. LDA is een voorbeeld van een onderwerp analysetechniek.



Hoe populair is tekstmining?

- > De grafiek hieronder laat het aantal relatieve wereldwijde zoekopdrachten voor “**Natural Language Processing**” bij Google zien; vooral tussen 2016 en 2018 is dit onderwerp heel populair geworden (en gebleven).
- > Het Marketing Science Instituut heeft als een van hun top onderzoeksprioriteiten voor 2018-2020: “*What approaches exist to capture and analyze nonstructured data such as (...) text in order to improve firm communications and the customer experience?*”



- > <https://www.msi.org/research/2018-2020-research-priorities/Capturing-Information-to-Fuel-Growth/4.4-approaches-to-ingesting-and-analyzing-data-to-drive-marketing-insights/>



Waarom is tekstmining belangrijk voor managers?

- › Bedrijven kampen met een groot aantal toenemende (klanten) data, waarvan veel ongestructureerde, geschreven data; klanten e-mails, reacties op sociale media, reviews, open antwoorden in tevredenheid enquêtes, etc.
- › Daarnaast proberen bedrijven steeds meer op klanten te focussen, hiervoor moeten zij de benodigheden, wensen en gedachten van hun klanten begrijpen en structureel analyseren.
- › Tekstmining kan helpen met het verbinden van deze twee uitdagingen. Met de juiste procedure kunnen managers een dashboard creëren met een 360 graden overzicht van hun klanten.
- › In dit rapport worden tekstmining methodes besproken welke kunnen helpen met het verkrijgen van inzichten van geschreven klantfeedback (en andere geschreven data). Voorbeelden worden gegeven over hoe bedrijven deze data kunnen gebruiken en hiermee kunnen verbeteren.



Overzicht van (nieuwe) tekstmining methodes

Onderzoek over tekstmining toont aan dat:

- › Er veel tools beschikbaar zijn voor tekstmining. Deze zijn waardevol voor marketingonderzoek en marketing uitvoering (Berger et al., 2019; Hartmann et al., 2019; Kübler et al., 2020)
- › Het sentiment van online geschreven opmerkingen op sociale media is zelfs een betere voorspelmethode dan (enquête gebaseerde) klanttevredenheid voor de toekomstig bedrijfsprestatie (bv. Marktaandeel, omzet, winst, abnormale aandelen omzet) (De Haan, 2020)
- › Het sentiment en het onderwerp van de geschreven feedback, bv. na klantcontact, kan helpen met het verbeteren van het voorspellen van toekomstig klantgedrag zoals churn. (De Haan en Menichelli, 2020)



Aanbevelingen voor marketing managers (1/2)

- › Er is veel waardevolle data beschikbaar als het gaat om geschreven feedback, bv. door klantcontact via sociale media. Deze data wordt over het algemeen niet systematisch geanalyseerd.
- › Om inzichten te verkrijgen raad ik aan:
 1. Selecteer de meest belangrijke geschreven bronnen welke gebruikt kunnen worden voor het verkrijgen van klantinzichten (bv. klantcontact, sociale media kanalen, review websites).
 2. Verzamel per periode (bv. dagelijks, wekelijks, maandelijks) de data van deze bronnen.
 3. Analyseer deze data systematisch met gebruik van sentiment of onderwerp analyse.



Aanbevelingen voor marketing managers (2/2)

- › Om inzichten te verkrijgen, raad ik aan (vervolgd):
 4. Link de resultaten aan een of meer KPI's (bv. retentie, marktaandeel, winst)
 5. Houd in de gaten en verbeter de prestatie van de klantendatabase door het gebruiken van de inzichten uit de geschreven data; het verkrijgen van inzichten is niet een doel op zichzelf. Het doel is om de waarde van en naar de klantendatabase te verbeteren!
- › In dit rapport worden praktische en theoretische voorbeelden gegeven, inclusief de (software) tools om dit in de praktijk uit te voeren.