Abstract

Introduction

De aanleiding voor ons mini-project is het artikel ‘Text analysis of Trump’s tweets confirms he writes only the (angrier) Android half’ in Variance Explained van David Robinson (Robinson, 2016) In dit artikel uit 2016 analyseert Robinson de tweets van Trump, omdat hem opvalt dat Trump vanaf een iPhone én vanaf een Samsung twittert.

Bij zijn analyse maakt Robinson gebruik van R en past hij text mining en sentiment analyse toe. Sentiment analyse betreft de analyse of een bericht positief, of negatief wordt gebruikt. Sentiment analyse wordt bijvoorbeeld gebruikt om te weten te komen hoe mensen over een bepaald product denken, of om te peilen hoeveel steun een politieke partij heeft.

De analyse kan plaatsvinden op basis van onderscheid tussen positieve en negatieve emoticons (Pak & Paroubek, 2010), of positieve, of negatieve woorden (Agarwal et al., 2011).

Met behulp van text mining laat Robinson zien dat bijna alle tweets van de Android quotes bevatten, dat de iPhone veel meer gebruik maakt van #, foto’s en links, dat de iPhone meer woorden als ‘join’ en ‘tomorrow’ gebruikt, terwijl meer emotioneel beladen woorden zoals ‘badly’, ‘crazy’, ‘weak’ en ‘dumb’ overtuigend van de Android komen.

Voor zijn sentiment analyse gebruikt Robinson tidytext package, dat woorden indeelt in 10 sentimenten: ‘positive’, ‘negative’, ‘anger’, ‘anticipation’, ‘disgust’, ‘fear’, ‘joy’, ‘sadness’, ‘surprise’ en ‘trust’. Hiermee laat hij zien dat het Android account 40-80% meer woorden gebruikt die aan ‘disgust’, ‘sadness’, ‘fear’, ‘anger’, en andere negatieve sentimenten gerelateerd zijn.

Robinson concludeert dat de Android en iPhone tweets duidelijk van verschillende mensen zijn, op verschillende tijden van de dag worden gepost en op verschillende manieren gebruik maken van #, links, foto’s en retweets. Ook concludeert hij dat de Android tweets bozer en negatiever zijn en hij vermoed daarom dat de Android van Trump zelf is en de iPhone van z’n campagne team.

Een jaar na dit onderzoek heeft Robinson (Robinson, 2017) opnieuw een analyse van het Twitteraccount @realDonaldTrump gedaan. Hierbij heeft hij nieuwe vragen gesteld en is op zoek gegaan naar specifieke woorden, zoals “Obama’, ‘Hillary’ en ‘fake’, als in ‘fake news’ en vooral naar wanneer Trump deze woorden is gaan gebruiken. 2016 was immers campagnetijd. Daarvoor deed Trump nog niet aan politiek en daarna was hij president. Robinson breidt de periode van de te onderzoeken tweets voor zijn onderzoek van 2017 uit van 2009 tot in 2017 en maakt een overzicht van de apparaten die voor het account @realDonaldTrump gebruikt worden en ontdekt dat Trump in maart 2017 van een Android is overgestapt naar een iPhone.

Deze onderzoeken van David Robinson hebben ons geïnspireerd om weer een jaar later opnieuw een sentiment analyse te doen op de tweets van @realDonaldTrump. In het afgelopen jaar is er namelijk weer veel veranderd. Zo is het maximum aantal karakters van een tweet uitgebreid van 140 naar 280. Argwal et al (2011) concludeert dat de resultaten van sentiment analyse in Twitter niet verschilt van analyses in langere teksten, terwijl Mayo (2014) juist wel een relatie aantoont tussen de lengte en het sentiment van tweets.

Dit alles gelezen hebbende heeft ons de volgende onderzoeksvraag doen stellen:

‘In hoeverre zijn er verschillen tussen de tweets van @realDonaldTrump in 2016 en 2018 van de Android en iPhone?

Related work

Robinson, D. (2016). Text analysis of Trump’s tweets confirms he writes only the (angrier) Android half. *Variance Explained*, *9*.

<http://varianceexplained.org/r/trump-tweets/>

Robinson, D. (2017). Trump's Android and iPhone tweets, one year later. *Variance Explained*.

<http://varianceexplained.org/r/trump-followup/>

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011, June). Sentiment analysis of twitter data. In *Proceedings of the workshop on languages in social media* (pp. 30-38). Association for Computational Linguistics.

Martínez-Cámara, E., Martín-Valdivia, M. T., Urena-López, L. A., & Montejo-Ráez, A. R. (2014). Sentiment analysis in Twitter. *Natural Language Engineering*, *20*(1), 1-28.

Mayo, M. (2014). A Clustering Analysis of Tweet Length and its Relation to Sentiment. *arXiv preprint arXiv:1406.3287*.

Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10, No. 2010).