

## The language of food security: an emotion and sentiment analysis of food security social media data

A. Molenaar<sup>1</sup>, D. Lukose<sup>2</sup>, L. Brennan<sup>3</sup>, E.L. Jenkins<sup>1</sup> and T.A. McCaffrey<sup>1</sup>

<sup>1</sup>Department of Nutrition, Dietetics & Food, Monash University, Notting Hill, Vic., Australia,

<sup>2</sup>Monash Data Futures Institute, Monash University, Clayton, Vic., Australia and

<sup>3</sup>School of Media & Communication, RMIT University, Melbourne, Vic., Australia

The use of data science to explore human language is an emerging area in health and nutrition and has so far, predominantly focussed on medical and disease related applications.<sup>(1,2)</sup> Data science techniques such as sentiment and emotion analysis can analyse large bodies of text to explore meaning behind human natural language.<sup>(3)</sup> This has applications in exploring the breadth of social media data around public health areas, including the topic of interest of this research, food security. This study aimed at exploring food security related social media data using different data science techniques, sentiment and emotion analysis. Twitter data from Australia containing key terms and hashtags relating to food security was collected using the Twitter API across 2019, 2020 and 2021. Data science techniques were used to analyse the language of the tweets, including sentiment (Very negative, Negative, Neutral, Positive and Very positive) using Valence Aware Dictionary and Sentiment Reasoner (VADER)<sup>(4)</sup> and emotion using two different emotion engines (six emotions: anger, fear, joy, love, sadness and surprise). A total of 107,897 food security-related tweets were collected across the 3 years, with 42.1% from 2020, followed by 31.0% from 2021 and 26.9% from 2019. The majority of tweets were re-tweets (69.6%), followed by replies (14.3%), original tweets (11.2%) and quote tweets (4.8%). Overall, the majority of tweets were classified as having a positive sentiment (47.6%), with 23.7% negative and 21.9% neutral. In the original emotion engine, the majority of tweets were classified as 'joy' (57.6%) followed by 'fear' (14.8%) and 'sadness' (13.7%). In the updated emotion engine similarly, the predominant emotion was 'joy' (43.7%), but the other emotions were evenly distributed, with 'love' being the second most frequent (17.6%), followed by 'sadness' (14.1%). The average proportions of sentiment and emotion categories were similar across the years; however, there were differences when disaggregated into months. There were peaks in negative sentiment in August 2019, December 2020 and October 2021. There was a peak in 'sadness' in January 2020 in both emotion engines, a peak of 'fear' in September 2019 in the original emotion engine and a peak of 'anger' in December 2019 and 'love' in March 2020 in the new emotion engine. Further analysis of the content of the tweets at these times will provide more context around the emotions and sentiment. Data science techniques may lack some context, especially in areas such as public health and food security which have not been used to develop the emotion and sentiment categorisation systems. There is a need for public health experts to be involved in the development and implementation of these data science techniques for them to successfully classify human language in public health areas such as food security.

### References

1. Gohil S, Vuik S & Darzi A (2018) *JMIR Public Health Surveill* 4, e43.
2. Zunic A, Corcoran P & Spasic I (2020) *JMIR Med Inform* 8, e16023.
3. Liu B (2012) *Synth Lect Hum Lang Technol* 5, 1–167.
4. Hutto CJ & Gilbert E (2014) *Proceedings Int AAAI Conf Web Soc Media* 8 (1), 216–225.