



INTISARI

Latar belakang : Obesitas telah menjadi masalah kesehatan yang dihadapi oleh negara-negara di dunia termasuk Indonesia karena prevalensinya yang terus meningkat dari waktu ke waktu. Berkaitan dengan obesitas, media sosial merupakan salah satu sarana yang digunakan oleh netizen untuk mengakses, berbagi dan berdiskusi seputar isu-isu obesitas. Twitter sebagai salah satu media sosial merupakan platform yang secara *real time* sering dipilih untuk mengkomunikasikan hal tersebut. Melalui analisis frekuensi kemunculan kata (*term frequencies*) dapat ditemukan *term* apa saja yang biasanya digunakan untuk mengkomunikasikan obesitas di Twitter selain juga pengelompokannya. Sedangkan analisis sentimen dapat digunakan untuk memahami bagaimana orang menggambarkan dan mengungkapkan persepsi mereka terhadap kondisi obesitas baik secara positif, negatif maupun netral. Kedua analisis tersebut penting untuk melihat sejauh mana media sosial seperti Twitter digunakan saat ini sebagai salah satu determinan diseminasi informasi kesehatan di Indonesia.

Tujuan Penelitian : Mengeksplorasi penggunaan data Twitter terhadap obesitas di Indonesia.

Metode penelitian : Penelitian ini termasuk jenis penelitian dengan rancangan *cross sectional*. Rancangan ini dipilih karena data yang diambil dari Twitter dalam seri waktu 5 tahun terakhir (2012 -2017). Hasil *webscraping* pada Twitter didapatkan jumlah tweet 67.942 tweet kemudian dicleaning dan menghasilkan sampel sebanyak 43.436 data tweet berbahasa Indonesia. Pengambilan data menggunakan paket ekstensi Google Chrome Twitter Testing v.01 dan dianalisis pada fase Pra-Proses menggunakan Python 3.7.2 dan R. Studio 3.5.2 untuk analisis dan visualisasi hasil.

Hasil Penelitian : Hasil analisis *term frequencies* terdapat 4.050 kata yang muncul dengan 10 (sepuluh) teratas dari kata tersebut yaitu “gemuk” (33.851), “gendut” (10.159), “makan” (8.721), “orang” (5.737), “status” (2.806), “obesitas” (2.799), “badan” (2.450), “perempuan” (2.189), “suka” (1.908) dan “kena” (1.544). Analisis sentimen tweet terkait obesitas menunjukkan bahwa sentimen positif merupakan yang terbanyak yaitu sekitar 22.246 tweet, diikuti oleh sentimen negatif sebanyak 12.015 tweet dan sentimen netral dengan jumlah 9.174 tweet dengan nilai akurasi sebesar 94% dan dikategorikan sebagai *Excellent Classification*.

Kesimpulan : Tingginya sentimen positif dibanding negatif dan netral menunjukkan bahwa ada kecenderungan tweet yang terkait obesitas berisi tweet-tweet positif sehingga perlu terus dipertahankan dan ditingkatkan melalui kampanye-kampanye kesehatan yang berkesinambungan. Selain itu diperlukan pengembangan sistem baru untuk analisis sentimen dengan metadata dari API Twitter yang lebih lengkap.

Kata kunci : Twitter, Analisis Sentimen, *Term Frequencies*, *Obesitas*



ABSTRACT

Background: Obesity has become a health problem faced by countries in the world including Indonesia because its prevalence continues to increase from time to time. Regarding obesity, social media is one of the tools used by netizens to access, share and discuss issues of obesity. Twitter as one of the social media is a platform that is often chosen in real time to communicate this. Through analysis of the frequency of occurrence (term frequencies) can be found what terms are usually used to communicate obesity on Twitter as well as grouping. While sentiment analysis can be used to understand how people describe and express their perceptions of obesity conditions both positively, negatively and neutrally. Both analyzes are important to see the extent to which Twitter social media is used as a determinant of the distribution of health information in Indonesia.

Research Objectives: Exploring the use of Twitter data on obesity in Indonesia.

Research Method: This study included the type of research with a cross-sectional design. This design was chosen because of data taken from Twitter in the last 5-year time series (2012-2017). The results of web scraping on Twitter found that the number of tweets of 67,942 tweets was then checked and produced a sample of 43,436 data tweets in Indonesian. Data retrieval uses the Google Chrome Twitter Testing v.01 extension package and is analyzed in the Pre-Process phase using Python 3.7.2 and R. Studio 3.5.2.

Research Results: Results of the analysis of term frequencies there are 4,050 words that appear with the top 10 (ten) of the words namely “*gemuk*” (33.851), “*gendut*” (10.159), “*makan*” (8.721), “*orang*” (5.737), “*status*” (2.806), “*obesitas*” (2.799), “*badan*” (2.450), “*perempuan*” (2.189), “*suka*” (1.908) and “*kena*” (1.544). Analysis of tweets related to obesity showed that the positive sentiment was the most, which was around 22,246 tweets. At the same time negative sentiments of 12,015 tweets and neutral sentiments with 9,174 tweets with an accuracy value of 94% and categorized as Excellent Classification.

Conclusion: The high positive sentiment shows there is a tendency for obesity-related tweets to contain positive tweets, that it needs to be maintained and improved through continuous health campaigns. In addition, it is necessary to develop a new system for sentiment analysis with more complete metadata from the Twitter API.

Keywords: Twitter, Sentiment Analysis, Term Frequencies, Obesitas