

Metode *Term Frequencies* untuk penelitian kesehatan di Twitter: Studi pada tweet berbahasa Indonesia terkait Obesitas

Felix F. Mailoa

Program Studi Ilmu Kesehatan Masyarakat, STIKES Papua Sorong

AIM / OBJECTIVE

Penggunaan media sosial seperti Twitter untuk penelitian kesehatan saat ini telah mengalami peningkatan yang cukup masif seiring dengan semakin terbukanya akses melalui ketersediaan API Twitter. Akses data yang mencapai 1%, data di Twitter telah memberikan manfaat untuk analisis baik melalui content analysis maupun studi prediksi dengan penerapan algoritma baru.

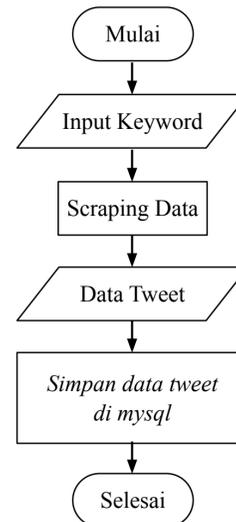
Namun di Indonesia, bidang penelitian ini belum menjadi pilihan utama bagi peneliti bidang informatika kesehatan masyarakat. Selain karena membutuhkan kemampuan komputasi juga karena perijinan akses API sebagai *developer* yang semakin selektif. Hal inilah yang menyebabkan terbatasnya jumlah riset kesehatan di Twitter khususnya tweet bahasa Indonesia. Imbasnya, tidak banyak tersedia rujukan penelitian khususnya terkait *keyword* untuk melakukan *query*.

Tujuan penelitian ini adalah untuk memberikan gambaran penggunaan metode *term frequencies* dalam penentuan *keyword* penelitian kesehatan masyarakat di Twitter.

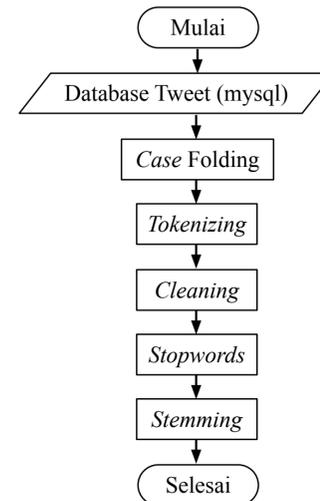
METHODS

Jenis penelitian adalah Deskriptif dengan menggunakan data Twitter selama 5 tahun (2012-2017) dari hasil *webscraping* menggunakan *keyword* "obesitas" OR "gemuk" OR "kegemukan" OR "gendut" OR "kegendutan" yang berjumlah 67.942 tweet. Kemudian dilakukan proses cleansing dan menghasilkan sampel sebanyak 43.436 data untuk fase pra-proses. Data dianalisis menggunakan Python 3.7.2 dan R Studio 3.5.2 untuk visualisasi *wordcloud*.

Pengambilan data tweet ini tidak melalui API *Search Twitter* namun menggunakan ekstensi *Twitter Testing v1.01* yang menggunakan *javascript* dan dipasangkan pada Google Chrome untuk melakukan *webscraping*. Seperti terlihat pada gambar *flowchart* berikut :



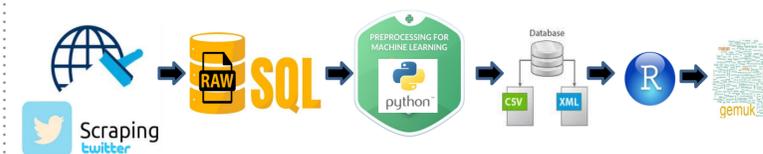
.Gambar 1 Alur webscraping



.Gambar 2 Praproses Tweet

Tweet yang telah masuk kemudian di lakukan pra-proses menggunakan Python. Hasil pra-proses disimpan dalam database baru untuk proses menggunakan R Studio.

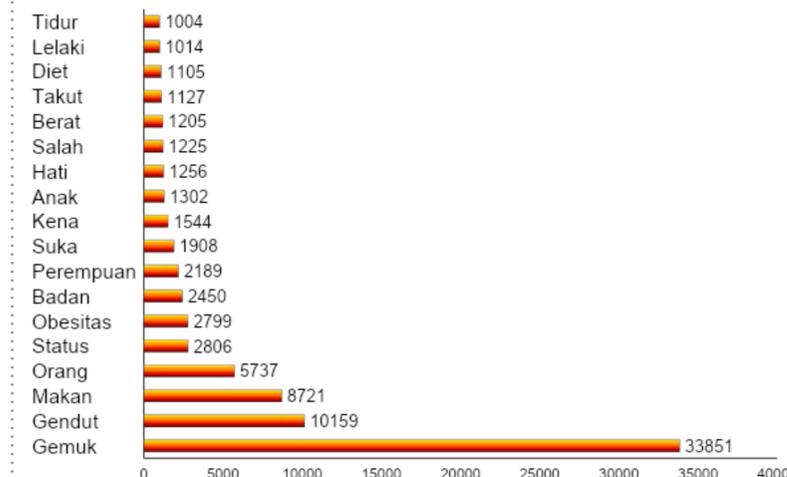
Library/Corpus kata yang digunakan untuk melakukan proses stemming bahasa Indonesia adalah dengan *Library Python Sastrawi*. Library ini merupakan pengembangan dari *library PHP sastrawi* dimana library tersebut menerapkan algoritma dari **Algoritma Nazief dan Adriani** (<https://devtrik.com/python/steeming-bahasa-indonesia-python-sastrawi/>).



.Gambar 3 Proses Scraping sampai Visualisasi Tweet

RESULTS

Frekuensi kata terkait obesitas dalam tweet yang berhasil dihimpun didapatkan dari proses perhitungan tingkat kemunculan kata dalam tiap tweet untuk perbincangan obesitas. *Term* yang berhasil terindeks sebanyak 4.050 kata. Hasil pengukuran frekuensi melalui R Studio adalah sebagai berikut :



.Gambar 1 Hasil Indeks *Term Frequencies*

Berdasarkan tingkat kemunculan kata dan visualisasi *wordcloud* (gambar 4), netizen cenderung menggunakan kata yang relevan dengan bahasa Indonesia ("gemuk" atau "gendut") karena preferensi bahasa yang digunakan dalam tweet tersebut merupakan asli bahasa Indonesia dan lebih umum digunakan dibandingkan kata "obesitas" yang merupakan kata dari serapan bahasa asing..



.Gambar 4 Visualisasi *Wordcloud*

CONCLUSIONS

Hasil penelitian ini menunjukkan bahwa dengan metode *Term Frequencies*, penggunaan kata/istilah di Twitter terkait obesitas diwakili oleh 4.050 kata dengan urutan 18 (delapan belas) teratas meliputi kata: "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), "kena" (1.544), "anak" (1.302), "hati" (1.256), "salah" (1.225), "berat" (1.205), "takut" (1.127), "diet" (1.105), "lelaki" (1.014) dan "tidur" (1.004).

Temuan dalam penelitian ini juga menunjukkan bahwa Twitter masih menyimpan potensi untuk diteliti berkaitan dengan kesehatan di Indonesia bahkan dengan metode yang sederhana. Penelitian ini memberikan gambaran cara orang berkomunikasi di media sosial tentang masalah kesehatan utamanya dengan penggunaan istilah. Kedepannya dari istilah-istilah yang terindeks dalam *Term Frequencies* ini bisa dikembangkan menjadi *keyword* baru. Sehingga kesulitan dalam menemukan *keyword* melalui jurnal dapat teratasi. Sangat disarankan untuk melakukan 2 (dua) kali penggunaan *keyword* untuk mendapatkan tweet yang betul-betul dibutuhkan.

Penelitian selanjutnya dapat mengeksplorasi data twitter dengan *content analysis*, *sentiment analysis*, *social network analysis* (SNA) dan model prediksi menggunakan data kasus (Twitter trends).

BIBLIOGRAPHY

- Chou, W.-Y. S., Prestin, A. and Kunath, S. (2014) 'Obesity in social media: a mixed methods analysis.', *Translational behavioral medicine*, 4(3), pp. 314–23. doi: 10.1007/s13142-014-0256-1.
- Eysenbach, G. (2009b) 'Infodemiology and Infovellance. Internet-based syndromic surveillance', in *OAHPP Syndromic Surveillance Workshop*. Ontario: Public Health Ontario (PHO) website.
- Harris, J. K. et al. (2014) 'Communication about childhood obesity on twitter', *American Journal of Public Health*, 104(7), pp. 62–69. doi: 10.2105/AJPH.2013.301860.
- Kim, A. R., Park, H. A. and Song, T. M. (2016) 'Development and Evaluation of an Ontology for Analyzing Social Data', *Stud Health Technol Inform*, 225(3), pp. 442–446. doi: 10.3233/978-1-61499-658-3-442.
- So, J. et al. (2016) 'What Do People Like to "Share" About Obesity? A Content Analysis of Frequent Retweets About Obesity on Twitter.', *Health communication*. Routledge, 31(2), pp. 193–206. doi: 10.1080/10410236.2014.940675.