

Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia

Felix Fridom Mailo¹, Lutfan Lazuardi²

¹Departemen Manajemen dan kebijakan Kesehatan Fakultas Kedokteran, Kesehatan Masyarakat dan Keperawatan Universitas Gadjah Mada.

²Departemen Sistem Informasi Manajemen Kesehatan Fakultas Kedokteran, Kesehatan Masyarakat dan Keperawatan, Universitas Gadjah Mada.

¹fridommailoa@gmail.com, ²lutfan.lazuardi@ugm.ac.id

ABSTRACT

Latar Belakang: 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 sentimen dengan metode text mining di Twitter, kita dapat memahami bagaimana orang menggambarkan dan mengungkapkan persepsi mereka terhadap kondisi obesitas baik secara positif, negatif maupun netral. Analisis tersebut penting untuk melihat sejauh mana media sosial seperti Twitter digunakan saat ini sebagai salah satu instrumen diseminasi informasi kesehatan di Indonesia. Tujuan penelitian untuk mengidentifikasi analisis sentimen pada Twitter terkait obesitas di Indonesia menggunakan metode text mining.

Metode penelitian: Jenis penelitian adalah cross sectional. Rancangan ini dipilih karena data yang diambil dari Twitter dalam seri waktu 5 tahun terakhir (2012 - 2017). Hasil webscraping pada Twitter didapatkan 67.942 tweet kemudian dicleaning dan menghasilkan sampel 43.436 data tweet berbahasa Indonesia. Pengambilan data menggunakan paket ekstensi Google Chrome Twitter Testing v.01 dan dianalisis menggunakan Python 3.7.2 dan R. Studio 3.5.2.

Hasil Penelitian: Berdasarkan hasil analisis sentimen tweet didapatkan sentimen positif sebanyak 22.246 (51,2%) tweet, diikuti oleh sentimen negatif sebanyak 12.015 (27,7%) tweet dan sentimen netral dengan jumlah 9.174 (21,1%) dari total 43.435 tweet. Nilai akurasi dengan algoritma Naïve Bayes didapatkan hasil nilai akurasi sebesar 94%.

Kesimpulan: Analisis sentimen tweet terkait obesitas dengan metode text mining lebih didominasi oleh sentimen positif dibanding sentimen negatif dan sentimen netral. Nilai akurasi dengan algoritma Naïve Bayes Classifier berada dalam kategori "Excellent Classification" yang artinya algoritma Naive Bayes Classifier berhasil memprediksi kategori sentimen dalam penelitian ini dengan baik. Analisis data selanjutnya dapat menggunakan data langsung dari API Twitter kemudian meningkatkan jumlah

kata kunci (keyword) dan menggunakan algoritma klasifikasi teks lainnya (SVM, DBN dll). Selain itu, dari sisi kesehatan masyarakat perlu upaya peningkatan literasi kesehatan melalui media sosial twitter terkait obesitas untuk meningkatkan pemahaman serta kemampuan mempercayai informasi kesehatan khususnya oleh organisasi kesehatan pemerintah guna mempertahankan sentimen positif publik yang sudah ada

Kata kunci: Analisis Sentimen, Obesitas, Text Mining, Twitter

ABSTRACT

Background: 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 sentiment analysis with the text mining method on Twitter, we can understand how people describe and express their perceptions of obesity conditions both positively, negatively and neutrally. The analysis is important to see the extent to which social media such as Twitter is used today as an instrument for disseminating health information in Indonesia. The study aimed to identify sentiment analysis on Twitter related to obesity in Indonesia using the text mining method.

Methods: The type of research is cross-sectional. 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 obtained 67,942 tweets then checked and produced a sample of 43,436 data tweets in Indonesian. Retrieval of data using the Google Chrome Twitter Testing v.01 extension package and analyzed using Python 3.7.2 and R. Studio 3.5.2.

Results: Based on the results of the tweet sentiment analysis, positive sentiments were 22,246 (51.2%) tweets, followed by negative sentiments of 12,015 (27.7%) tweets and neutral sentiments 9,174 (21.1%) of the total 43,435 tweets. The value of accuracy with the Naïve Bayes algorithm results in an accuracy value of 94%.

Conclusions: *Positive sentiments dominated the analysis of the sentiment of tweets related to obesity with the text mining method compared to negative sentiments and neutral sentiments. The accuracy value with the Naïve Bayes Classifier algorithm is in the category of "Excellent Classification". Which means that the Naive Bayes Classifier algorithm successfully predicts the sentiment category in this research well. Further data analysis can use data directly from the Twitter API then increase the number of keywords and use other text classification algorithms (SVM, DBN etc.). Also, in terms of public health need efforts to improve health literacy through Twitter about obesity. Improve understanding and ability to trust health information, especially by government health organizations to maintain the positive sentiment that already exists.*

Keywords: *Sentiment Analysis, Obesity, Text Mining, Twitte*

PENDAHULUAN

Obesitas telah menjadi masalah kesehatan yang dihadapi oleh negara-negara di dunia, bukan hanya negara maju tetapi juga negara berkembang terutama di daerah perkotaan karena prevalensinya yang terus meningkat dari waktu ke waktu. Menurut data¹ kelebihan berat badan dan obesitas telah menunjukkan peningkatan yang nyata selama 4 dekade terakhir. Tingkat obesitas di tahun 1975 pada pria sekitar 3% dan pada wanita di atas 6% sementara kelebihan berat badan selama periode waktu yang sama di bawah 21% pada pria dan hanya di bawah 23% pada wanita. Di tahun 2014, 39% pria dan 40% wanita berusia lebih dari 18 tahun kelebihan berat badan dan dari jumlah tersebut 11% pria dan 15% wanita mengalami obesitas. Sehingga dengan demikian hampir 2 miliar orang dewasa di seluruh dunia kelebihan berat badan. Berdasarkan jumlah ini lebih dari setengah miliar orang mengalami obesitas.

Di Indonesia prevalensi kasus obesitas umum (digunakan untuk kategori berat badan lebih dan obesitas) pada tahun 2007 untuk penduduk umur ≥ 15 tahun sebesar 19,1% (8,8% berat badan lebih dan 10,3% obesitas)². Kemudian di tahun 2013 tren itu meningkat, prevalensi gemuk pada kelompok umur >18 tahun sebanyak 28,9% (13,5% berat badan lebih dan 15,4% obesitas) dan untuk remaja umur 16-18 tahun sebanyak 7,3% (5,7% gemuk

dan 1,6% obesitas)³. Hanya dalam kurun waktu 3 tahun (2010-2013) prevalensi gemuk pada kelompok umur 16-18 tahun naik dari 1,4% menjadi 7,3% dan kelompok umur >18 tahun naik dari 21,6% menjadi 28,9%.

Terkait obesitas, media sosial sangat erat hubungannya dengan perilaku pencarian informasi yang berkaitan dengan suatu fenomena kesehatan (*health seeking information*). Peranan media sosial sebagai solusi untuk melawan obesitas dapat dilihat melalui dukungan melawan stigma⁴ atau digunakan untuk berbagi informasi kesehatan online⁵ secara gratis⁶ karena dapat dirahasiakan dan kredibel⁷. Salah satu media sosial yang dapat digunakan untuk analisa data adalah Twitter.

Data tweet adalah contoh sumber data *real-time* dunia nyata tentang perilaku kesehatan yang mungkin mewakili konten yang lebih naturalistik dalam volume besar⁸. Beberapa analisis data Twitter menunjukkan adanya kecenderungan tweet terkait obesitas berbicara tentang penurunan berat badan⁹, praktik diet individual dan kurangnya aktivitas fisik¹⁰, bahkan penyebarannya secara spasial¹¹. Selain itu twitter sebagai sarana komunikasi untuk makanan dan analisis risikonya terhadap kesehatan^{12,13,14} Obesitas pada sosial media juga berkaitan dengan stigma berat badan dan isolasi sosialnya^{15,16} bahkan telah sampai pada tingkat pencegahannya¹⁷.

Beberapa pendekatan yang dapat dilakukan misalnya metode *text mining*¹⁸ dengan pendekatan analisis sentimen publik (netizen). Pendekatan analisis sentimen bertujuan untuk memahami bagaimana orang menggambarkan dan mengungkapkan perasaan mereka terhadap kondisi obesitas baik secara positif, negatif maupun netral. *Sentiment analysis* atau *opinion mining* adalah salah satu cabang dari *text classification* yang mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik, dan *text mining* bertujuan menganalisis pendapat, sentimen, evaluasi, sikap, penilaian, dan emosi seseorang pembicara atau penulis berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu lainnya¹⁹.

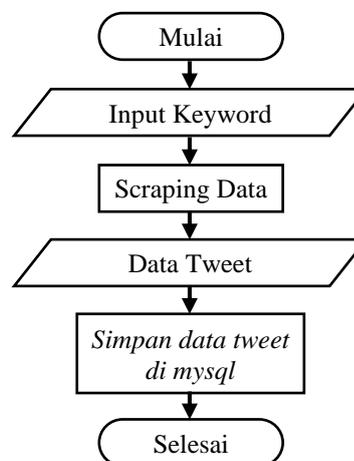
Tugas dasar dalam analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen, kemudian menentukan pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif atau negatif. Sentimen analisis juga dapat menyatakan perasaan emosional sedih, gembira, atau marah. Ekspresi atau *sentiment* mengacu pada fokus topik tertentu, pernyataan pada suatu topik mungkin akan berbeda makna dengan pernyataan yang sama pada *subject* yang berbeda. Sentimen analisis bertujuan menganalisis pendapat, sentimen, evaluasi, sikap, penilaian, dan emosi seseorang pembicara atau penulis berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu lainnya

Berdasarkan data dari Asosiasi Penyedia Jasa Internet Indonesia (APJII) saat ini di Indonesia dari total jumlah penduduk sebanyak 256,2 juta jiwa, 132,7 juta diantaranya adalah pengguna internet. Jumlah ini mengalami kenaikan dari tahun 2014 yaitu sebanyak 88,1 juta. Dari 132,7 juta jiwa pengguna internet tersebut 97,4% (129,2 juta) adalah pengguna media sosial dan dari jumlah tersebut 5,5% (7,2 juta) adalah pengguna aktif Twitter dengan rata-rata 13,8% pengguna (18,3 juta) menggunakan internet untuk mencari informasi seputar masalah kesehatan²⁰. Melihat potensi tersebut seharusnya dapat dimaksimalkan namun sampai saat ini belum banyak penelitian yang dilakukan.

METODE PENELITIAN

Rancangan penelitian ini adalah *cross sectional*. Tujuan penelitian untuk mengidentifikasi analisis sentimen pada Twitter terkait obesitas di Indonesia menggunakan metode *text mining*. Sampel pada penelitian ini adalah tweet netizen Indonesia yang terkait obesitas yaitu yang mengandung kata-kata kunci “obesitas” OR “gemuk” OR “kegemukan” OR “gendut” OR “kegendutan”¹⁵. Periode pengambilan data tweet dari 01 Januari 2012 sampai 31 Desember 2017. Hasil *webscraping* (gambar 1) pada Twitter didapatkan jumlah tweet 67.942 tweet berbahasa Indonesia kemudian dilakukan cleaning dan menghasilkan sampel sebanyak

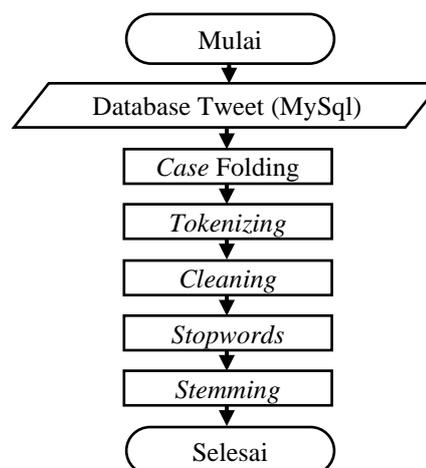
43.436. Terdapat beberapa instrumen/tool berupa aplikasi yang digunakan dalam penelitian ini yaitu Twitter Testing versi 1.0 (aplikasi webscraping), Python versi 3.7.2 dan R Studio versi 3.5.2.



Gambar 1. Alur Scraping Data

HASIL

Seperti halnya dalam *data mining*, aplikasi *text mining* pada suatu studi kasus, harus dilakukan sesuai prosedur analisis. Langkah awal sebelum suatu data teks dianalisis menggunakan metode-metode dalam *text mining* adalah melakukan *pre-processing* teks²¹. Database hasil ekstrak diolah melalui fase Praproses (Gambar 2) menggunakan aplikasi Python untuk menjalani serangkaian tindakan yang meliputi *Query* data MySQL, *Case Folding*, *Tokenizing*, *Cleaning*, *Stopword* dan *Stemming*.



Gambar 2. Alur Praproses Data

Dalam praproses ini 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 Nazief & Adriani²².

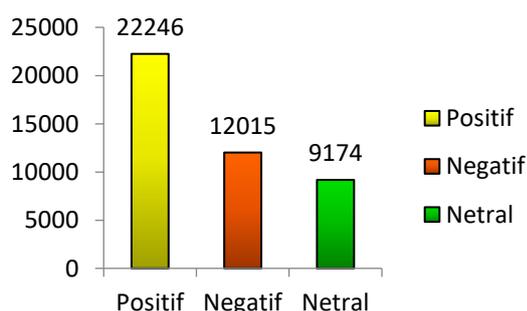
Bentuk tanggapan netizen di twitter terkait dengan obesitas dapat dilihat dalam bentuk analisis sentimen. Analisis sentimen merupakan gambaran tentang perasaan

yang ditangkap dari setiap kata dalam “cuitan” menjadi algoritma untuk kemudian diklasifikasikan sesuai dengan kamus kata (corpus). Sentimen publik/netizen dalam sosial media dibedakan atas sentimen positif, negatif dan juga netral. Berikut ini contoh tweet yang mengandung sentimen positif, negatif dan netral:

Tabel 1. Contoh Tweet

No	Akun	Tweet	Tipe Sentimen
1	@Ri**Bilqis	Yang sehat itu yang lucu dan bagus, bukan yang gemuk/obesitas #LawanDiabetes #HKS2016 @KemenkesRI	Positif
2	@rez**i	Terakhir hal yang mungkin sebenarnya lo tau tapi lo engga sadar. Penjual makanan, sejauh ini yang gue tau ga pernah peduli sama apa yang lo makan, lo mau obesitas kek atau penyakit jantung mereka ga peduli, yang mereka peduli adalah jualan mereka laku.	Negatif
3	@ok**onenews	Sering Diejek Gemuk Picu Seseorang Alami Obesitas #ExperienceTheZone http://okz.me/QcFI	Netral

Tugas dasar dalam analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen, kemudian menentukan pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif atau negatif. *Sentiment analysis* juga dapat menyatakan perasaan emosional sedih, gembira, atau marah. Secara keseluruhan, hasil analisis sentimen twitter berbahasa indonesia terkait obesitas dari tahun 2012 sampai dengan tahun 2017, dapat dilihat pada gambar 3 berikut ini :



Gambar 3. Hasil Analisis Sentimen Tweet terkait Obesitas di Indonesia

Berdasarkan hasil analisis sentimen tweet menunjukkan bahwa sentimen positif merupakan yang terbanyak yaitu sekitar 22.246 (51,2%) tweet, diikuti oleh sentimen negatif sebanyak 12.015 (27,7%) weet dan sentimen netral dengan jumlah 9.174 (21,1%) tweet dari

total 43.435 tweet. Untuk menghitung ketepatan penentuan kelas sentimen maka diperlukan perhitungan nilai akurasi.

Kinerja suatu algoritma dapat dilihat dari besar kecilnya nilai akurasi. Untuk menghitung ketepatan penentuan kelas sentimen maka diperlukan perhitungan nilai akurasi. Akurasi hasil analisis sentimen ini dihitung menggunakan algoritma NBC (*Naive Bayes Classifier*) dengan pertimbangan tertentu²³ untuk aplikasi dunia kesehatan²⁴.

Hasil akurasi terhadap suatu metode analisis sentimen sangat dipengaruhi oleh beberapa hal antara lain; 1) jumlah data *training* dan *testing*, 2) jumlah data set yang digunakan, dan 3) komposisi jumlah data positif dan negatif²⁵. Dalam uji akurasi, jumlah tweet yang digunakan sebanyak 34.261 tweet dari jumlah 43.435 tweet. Hal ini terjadi karena yang diikutkan masuk dalam uji akurasi hanya tweet yang memiliki sentimen positif dan sentimen negatif. Data tersebut dibagi 2 (dua) bagian yaitu masing-masing untuk data *testing* dan data *training*. Adapun data *training* partisinya sebanyak 75% yaitu dari range 1 – 25.695 sedangkan data *testing* partisinya sebanyak 25% yaitu dari range 25.696 – 34.261. Hasil uji akurasinya dapat dilihat pada tabel berikut ini:

Tabel 2. Hasil Uji Akurasi Data

Item Pengujian	Nilai
Akurasi	: 0.9398
95% CI	: (0.9345 – 0.9447)
Presisi	: 0.9431
Sensitifitas	: 0.8808
Spesifitas	: 0.9715
PPV	: 0.9381
NPV	: 0.9533

Dari tabel diatas, didapatkan hasil uji akurasi sebesar 94% dan berdasarkan nilai standart²⁶ dapat disimpulkan bahwa algoritma *Naive Bayes classifier* cukup berhasil memprediksi kategori sentimen karena dari hasil uji akurasi masuk dalam kategori “*Excellent Classification*”.

PEMBAHASAN

Analisis sentimen obesitas menggunakan metode *text mining* dalam penelitian ini mendapatkan hasil bahwa netizen di Twitter memiliki sentimen positif lebih dominan dibanding sentimen negatif dan netral. Hasil penelitian ini berbeda dengan penelitian yang dilakukan oleh ²⁷ yang mendapatkan hasil bahwa sentimen publik di sosial media cenderung lebih ke arah yang netral dibanding positif maupun negatif. Sentimen netral mendominasi sebanyak 66% kemudian sentimen positif sebanyak 22% dan negatif sebanyak 12%. Namun walaupun berbeda, salah satu temuan yang penting adalah bahwa selama rentang 3 tahun (2011-2013) persepsi/sentimen yang sifatnya positif mengalami peningkatan signifikan. Sentimen positif dan negatif mengalami penurunan rata-rata sebesar 1% per tahunnya sedangkan untuk sentimen positif mengalami kenaikan rata-rata 1 – 1,3% per tahunnya. Tren kenaikan ini menunjukkan bahwa mulai tumbuh kesadaran pada publik yang berkomunikasi tentang obesitas di sosial media akan pentingnya dukungan secara konkrit, misalnya melalui kampanye melawan stigma dibanding melakukan *cyberbullying* terhadap orang dengan obesitas. Hal ini juga menunjukkan bahwa media sosial tidak boleh dilihat hanya sebagai tempat berkembang biaknya stigma berat badan, tetapi juga lingkungan yang dapat melindungi individu yang kelebihan berat badan dari stigma¹⁵.

Pemilihan penggunaan metode *Naive Bayes* untuk klasifikasi lebih dipilih karena mempunyai kemampuan klasifikasi teks juga akurasi yang baik dibanding beberapa metode lainnya. Perhitungan klasifikasi dalam penelitiannya dengan menggunakan metode *Naive Bayes* menghasilkan nilai pengujian yang lebih baik secara keseluruhan dari pada menggunakan metode *Deep Belief Network* dan *Support Vector Machine*²⁸. Salah satu hal juga yang menyebabkan terjadi perbedaan dalam hasil penelitian ini adalah perbedaan bahasa serta corpus yang digunakan. Belum ditemukannya penelitian analisis sentimen yang terkait obesitas pada twitter berbahasa Indonesia sehingga dalam proses komparasi hasil, masih menggunakan tweet dari bahasa asing. Hal ini penting karena perbedaan term dan model klasifikasi ditentukan oleh preferensi bahasa yang juga akan berdampak pada penggunaan corpus dalam konsep *machine learning*. Diperlukan pengembangan corpus berbahasa Indonesia khusus bidang kesehatan untuk penelitian twitter di Indonesia. Ada kebutuhan untuk meningkatkan keakuratan analisis dengan menggunakan metode analisis sentimen yang dilatih dengan corpus kesehatan khusus (terkait layanan kesehatan) baik untuk perangkat lunak dikembangkan sendiri atau paket produk komersial²⁹.

Besarnya angka atau jumlah sentimen positif dibanding negatif dan netral bisa dimungkinkan berhubungan dengan banyaknya jumlah tweet yang positif baik untuk yang ada disektor individu maupun sektor non individu (pemerintah, media, lembaga profit maupun non profit dll). Mekanisme *text mining* dengan akurasi yang baik ini dapat membantu menemukan tanggapan dalam cuitan sehingga memungkinkan didapatkannya persepsi seseorang atau kelompok terhadap fenomena kesehatan tertentu. Seperti yang dinyatakan oleh Widener Li³⁰ dalam penelitian mereka bahwa semakin positif tweet itu, semakin besar kemungkinan tweet itu mengandung konten sehat. Sehat tidaknya suatu konten yang berpengaruh pada nilai dari sentimen yang dianalisis juga perlu dipahami secara terstruktur yakni dari kacamata komunikasi massa. Teknologi elektronik saat ini menyediakan lebih banyak

informasi dan layanan terhadap orang pada kecepatan yang lebih tinggi pada jarak yang lebih jauh. Sejumlah layanan bersifat interaktif meski kini berada pada level yang relatif rendah. Ini memungkinkan semakin banyak informasi yang beredar pada kecepatan yang lebih tinggi pada jarak yang lebih jauh³¹ termasuk informasi kesehatan.

Memahami obesitas pada sosial media seperti twitter adalah berkaitan erat dengan persepsi seseorang dalam memandang isu obesitas ini. Dari sisi komunikasi, stigma dan dukungan terhadap obesitas adalah dua hal yang bertolak belakang namun posisi moral dalam mendukung atau tidak mendukung salah satunya tergantung pada persepsi pribadi. Tingginya nilai sentimen positif dibanding sentimen negatif dan netral bisa dikaitkan dengan persepsi netizen yang cenderung positif terhadap isu obesitas. Suatu topik akan berkembang dalam Twitter secara masif tergantung ruang dan waktu serta dipengaruhi situasi sosial-ekonomi yang artinya saat ini persepsi obesitas di twitter masih dalam batasan yang wajar. Persepsi terhadap isu di media sosial akan didorong oleh kemampuan menginterpretasikan informasi, bukan interpretasi terhadap objek tersebut melainkan bagaimana tampak objek tersebut³². Untuk mampu melakukan interpretasi yang baik terhadap suatu fenomena di media sosial diperlukan pengetahuan yang cukup dan pengetahuan itu datang salah satunya dari literasi media sosial terkait kesehatan.

Keberadaan literasi media sosial akan mendukung upaya kritis dalam memahami sifat, teknik dan dampak dari pola komunikasi dalam media khususnya Twitter. Literasi media sosial tidak hanya bertumpu pada satu kutub negatif atau positif terhadap berbagai sentimen publik tetapi lebih dari pada itu literasi media akan membentuk pola otomatisasi berpikir netizen untuk membangun jiwa kritis terhadap isu-isu kesehatan di media sosial. Literasi media membantu melakukan normalisasi terhadap dampak negatif informasi serta komunikasi dengan tidak menyandarkannya hanya pada satu persepektif tetapi mencakup beragam perspektif. Literasi media tidak hanya merupakan sebuah kesadaran

dan pemahaman yang lebih baik, tetapi juga merupakan otonom yang sangat penting sebagai fungsi kontrol. Melalui literasi media sosial, seseorang diajak untuk berpikir kritis untuk memahami persepektif isu atau topik. Penggunaan media sosial secara kritis merupakan kunci penting literasi media sosial di Indonesia saat ini³³.

Literasi media sosial bukan upaya membatasi penggunaan bahkan akses media sosial tetapi untuk meningkatkan kesadaran dalam mengonsumsi media sosial secara lebih kritis dan terkendali. Praktisi kesehatan masyarakat dan penyedia layanan kesehatan yang berada pada level preventif harus menyadari sifat “percakapan online” tentang obesitas, bukan hanya memahami tetapi mampu terlibat untuk menjadi terdepan dalam kampanye-kampanye hidup sehat. Termasuk memahami sentimen negatif yang bisa menghilangkan pesan kesehatan serta sentimen positif yang mempromosikan kesehatan.

Organisasi kesehatan pemerintah juga perlu meningkatkan literasi media sosial terkait kesehatan agar masyarakat mampu memahami bagaimana menggunakan situs jejaring sosial sebagai alat komunikasi yang hemat biaya untuk menyampaikan informasi kesehatan yang penting. Lini penelitian ini juga dapat membantu organisasi kesehatan mendesain pesan kesehatan yang dirancang untuk audiens target, serta mengembangkan strategi komunikasi untuk audiens umum³⁴ Terutama potensi Twitter untuk membantu membentuk kemitraan dengan audiens dan melibatkan mereka sebagai peserta program kesehatan yang dapat mengarah pada tindakan untuk meningkatkan kesehatan³⁵.

Dominasi sentimen yang positif dari hasil *text mining* ini harus terus dipertahankan dan perlu lebih ditingkatkan lagi melalui kampanye-kampanye kesehatan dengan merujuk pada bukti-bukti yang ada. Misalnya mendukung kampanye terkait diet yang sehat, peningkatan frekuensi aktifitas fisik secara berarti, pengendalian/manajemen stress yang baik, menghindari *body shaming* dan *cyberbullying* dll. Hal ini dapat dilakukan melalui desain informasi yang lebih kekinian melalui pendekatan/strategi komunikasi yang lebih

mewadahi keberadaan kaum milenial sehingga lebih terarah dan tepat sasaran. Kesemua upaya ini untuk mengurangi ancaman stigma sosial yang kapan saja dapat terjadi. Merujuk pada beberapa kasus obesitas yang belakangan ini sering terjadi dan menjadi viral di media sosial, perlu kiranya kanal-kanal informasi yang baik untuk mengurangi kemungkinan terjadinya *hoax* informasi kesehatan yang seringkali berujung dengan iklan. *Hoax* terhadap informasi ini juga dapat menyebabkan *misinterpretasi* dan *mispresepsi* terhadap upaya-upaya penanganan obesitas di level individu.

Temuan penelitian ini merupakan salah satu dasar yang masih sangat mungkin dikembangkan kedepannya. Hal ini memberikan jalan terhadap semakin spesifiknya jenis topik kesehatan yang dapat diambil dengan akses ke dalam database Twitter (API). Sehingga mendorong analisis terhadap tren yang sedang terjadi berdasarkan tingkat percakapannya lalu melakukan validasi secara menyeluruh dengan data kesehatan yang sama di level nasional atau memadupadankan data pada media sosial dengan data pada mesin pencari seperti Google. Pendekatan-pendekatan ini bisa membantu tersedianya data awal yang lebih cepat guna memotong mata rantai penularan pada beberapa kasus penyakit menular yang mungkin saja *unpredictable*.

KESIMPULAN

Berdasarkan hasil analisis sentimen tweet dengan metode *text mining* didapatkan sentimen positif lebih dominan yaitu sebanyak 22.246 (51,2%) tweet, diikuti oleh sentimen negatif sebanyak 12.015 (27,7%) tweet dan sentimen netral dengan jumlah 9.174 (21,1%) dari total 43.435 tweet. Nilai akurasi dengan algoritma *Naive Bayes Classifier* didapatkan hasil sebesar 94% dan berada dalam kategori "*Excellent Classification*" yang artinya algoritma *Naive Bayes Classifier* berhasil memprediksi kategori sentimen dalam penelitian ini dengan baik.

Menurut penelliti ada beberapa hal yang perlu disarankan berdasarkan temuan yaitu analisis data berikutnya dapat menggunakan data langsung dari API Twitter kemudian meningkatkan jumlah kata kunci

(*keyword*). Serta menggunakan algoritma yang lain seperti *Support Vector Machines (SVM)*, *Deep Belief Network (DBN)*, dan algoritma klasifikasi teks lainnya. Selain itu dari sisi kesehatan masyarakat perlu upaya peningkatan literasi kesehatan melalui media sosial twitter terkait obesitas, untuk meningkatkan pemahaman serta kemampuan mempercayai informasi kesehatan khususnya oleh organisasi kesehatan pemerintah guna mempertahankan sentimen positif publik yang sudah ada.

DAFTAR PUSTAKA

1. WHO. Global Health Observatory (GHO) data (Overweight and obesity). http://www.who.int/gho/ncd/risk_factors/overweight_text/en/. Published 2016. Accessed May 5, 2017.
2. Depkes RI B. Laporan Nasional Riset Kesehatan Dasar Tahun 2007. Jakarta: Depkes, RI; 2008.
3. Depkes RI B. Riset Kesehatan Dasar 2013. Depkes, RI; 2013. doi:1 Desember 2013
4. Yoo JH, Kim J. Obesity in the New Media: A Content Analysis of Obesity Videos on YouTube. *Health Commun.* 2012;27(1):86-97. doi:10.1080/10410236.2011.569003
5. Hausmann JS, Touloumtzis C, White MT, Colbert JA, Gooding HC. Adolescent and Young Adult Use of Social Media for Health and Its Implications. *J Adolesc Heal.* 2017;XXX:1-6. doi:http://dx.doi.org/10.1016/j.jadohealth.2016.12.025
6. Ybarra ML, Suman M. Help seeking behavior and the Internet: A national survey. *Int J Med Inform.* 2006;75(1):29-41. doi:10.1016/j.ijmedinf.2005.07.029
7. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: The place of the internet. *Soc Sci Med.* 2005;60(7):1467-1478. doi:10.1016/j.socscimed.2004.08.010
8. Yoon S, Elhadad N, Bakken S. A practical approach for content mining of tweets. *Am J Prev Med.* 2013;45(1):122-129. doi:10.1016/j.amepre.2013.02.025
9. Pagoto S, Schneider KL, Evans M, et al. Tweeting it off: characteristics of adults who tweet about a weight loss attempt. *J Am Med Inform Assoc.*

- 2014;21:1032-1037. doi:10.1136/amiajnl-2014-002652
10. So J, Prestin A, Lee L, Wang Y, Yen J, Chou W-YS. What Do People Like to "Share" About Obesity? A Content Analysis of Frequent Retweets About Obesity on Twitter. *Health Commun.* 2016;31(2):193-206. doi:10.1080/10410236.2014.940675
 11. Gore RJ, Diallo S, Padilla J. You are what you tweet: Connecting the geographic variation in America's obesity rate to twitter content. *PLoS One.* 2015;10(9):1-16. doi:10.1371/journal.pone.0133505
 12. Shan LC, Panagiotopoulos P, Regan Á, et al. Interactive communication with the public: Qualitative exploration of the use of social media by food and health organizations. *J Nutr Educ Behav.* 2015;47(1):104-108. doi:10.1016/j.jneb.2014.09.004
 13. Rutsaert P, Regan Á, Pieniak Z, et al. The use of social media in food risk and benefit communication. *Trends Food Sci Technol.* 2013;30(1):84-91. doi:10.1016/j.tifs.2012.10.006
 14. Kuttschreuter M, Rutsaert P, Hilverda F, Regan Á, Barnett J, Verbeke W. Seeking information about food-related risks: The contribution of social media. *Food Qual Prefer.* 2014;37:10-18. doi:10.1016/j.foodqual.2014.04.006
 15. Chou W-YS, Prestin A, Kunath S. Obesity in social media: a mixed methods analysis. *Transl Behav Med.* 2014;4(3):314-323. doi:10.1007/s13142-014-0256-1
 16. Primack BA, Shensa A, Sidani JE, et al. Social Media Use and Perceived Social Isolation Among Young Adults in the U.S. *Am J Prev Med.* 2017;53(1):1-8. doi:10.1016/j.amepre.2017.01.010
 17. Park BK, Nahm E-S, Rogers VE, et al. A Facebook-Based Obesity Prevention Program for Korean American Adolescents: Usability Evaluation. *J Pediatr Heal Care.* 2016;31(1):1-10. doi:10.1016/j.pedhc.2016.02.002
 18. Amarullah A. Menganalisis Data dari Twitter. http://www.academia.edu/11823091/Menganalisis_Data_dari_Twitter. Published 2015. Accessed May 14, 2017.
 19. Liu S, Young SD. A survey of social media data analysis for physical activity surveillance. *J Forensic Leg Med.* 2016;1-4. doi:10.1016/j.jflm.2016.10.019
 20. Asosiasi Penyelenggara Jasa Internet Indonesia (APJII). *Statistik Pengguna & Perilaku Pengguna Internet Indonesia.* 2016. <http://www.apjii.or.id/survei2016>.
 21. Karimah PAS. Analisis Kompetitif Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes Classifier (NBC) Dan Support Vector Machine (SVM): Studi Kasus Di Industri Transportasi Umum Taksi. 2018.
 22. Nazief, Adriani. *Ssteeming Bahasa Indonesia Python Sastrawi.* devtrik.com. <https://devtrik.com/python/steeming-bahasa-indonesia-python-sastrawi/>. Published 2018. Accessed January 10, 2019.
 23. Sudiantoro AV, Zuliarso E, Studi P, et al. Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Classifier. In: *SINTAK 2018.* ; 2018:398-401.
 24. Poucke S Van, Zhang Z, Schmitz M, et al. Scalable Predictive Analysis in Critically Ill Patients Using a Visual Open Data Analysis Platform. *PLoS One.* 2016;1-21. doi:10.1371/journal.pone.0145791
 25. Novantirani A, T MKSS, Effendy V, et al. Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. In: *E-Proceeding of Engineering. Vol 2.* ; 2015:1177-1183.
 26. Gorunescu F. *Data Mining: Concepts, Models and Techniques.* Springer; 2011.
 27. Kim AR, Park HA, Song TM. Development and Evaluation of an Ontology for Analyzing Social Data. *Stud Heal Technol Inf.* 2016;225(3):442-446. doi:10.3233/978-1-61499-658-3-442
 28. Zulfa I, Winarko E. Sentimen Analisis Tweet Berbahasa Indonesia dengan Deep Belief Network. *IJCCS.* 2017;11(2):187-198.
 29. Gohil S, Vuik S, Darzi A. Sentiment Analysis of Health Care Tweets : Review of the Methods Used Corresponding Author : JMIR public Heal Surveill. 2018;4(2):1-9. doi:10.2196/publichealth.5789
 30. Widener MJ, Li W. Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Appl Geogr.* 2014;54:189-197. doi:10.1016/j.apgeog.2014.07.017
 31. Ayun PQ, Pratiwi MR, Boer KM, et al. *Cyberspace and Culture.* 1st ed. (Sukmono FG, ed.). Yogyakarta: Lembaga Penerbit Fakultas Ekonomi Universitas Indonesia; 2014.
 32. Mulyana D. *Ilmu Komunikasi, Suatu Pengantar.* 17th ed. (Muchlis, ed.). Bandung: PT. Remaja Rosdakarya;

2017.

33. Hermawan H. Literasi Media, Kesadaran Dan Analisis. 1st ed. Yogyakarta: Calpulis; 2017.
34. Park H, Rodgers S, Stemmler J. Analyzing Health Organizations' Use of Twitter for Promoting Health Literacy. *J Med Internet Res.* 2013;730. doi:10.1080/10810730.2012.727956
35. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Current BS, Reese H. Use of Twitter Among Local Health Departments : An Analysis of Information Sharing , Engagement , and Action Corresponding Author : *J Med Internet Res.* 2013;15. doi:10.2196/jmir.2775