

**TEXT MINING DAN SENTIMEN ANALISIS TWITTER PADA GERAKAN LGBT****Hartanto**¹✉¹ Fakultas Psikologi, Universitas Widya Dharma, Klaten**Info Artikel***Sejarah Artikel:*

Diterima 9 Januari 2017

Disetujui 20 Februari 2017

Dipublikasikan 1 Maret 2017

*Keywords:**text mining, sentiment analysis, machine learning, LGBT movement.***Abstrak**

Tujuan penelitian ini untuk mendeskripsikan frekuensi opini kicauan di Twitter terkait pro dan kontra terhadap gerakan LGBT. Riset ini termasuk dalam penelitian data mining kuantitatif dan terbagi dalam dua tahap. Pertama, analisis *text mining wordclouds* dan dilanjutkan analisis sentiment dalam tweet. Pendekatan sub metode *bags of word* dilakukan untuk merangkai sentiment dari twitter. *Negative* dan *positive words* murni melakukan perubahan dari *sentiment lexicon* Bing Liu dengan melakukan adaptasi ke bahasa Indonesia. *Stop words* yang merupakan bagian dari analisis *text mining* juga disesuaikan untuk kebutuhan analisis. Hasil yang didapat mengenai gerakan LGBT cukup konsisten dengan kondisi terkini, hasil dari wordcloud simetris dengan histogram analisis sentiment, dimana frekuensi kata seperti “kumpul kebo”, “moral”, “Indonesia”, “perbincangan”, “dibui”, “pindah”, “ditantang”, “garis batas” terbilang tinggi, dan sebanyak 379 tweet beropini netral, 79 menyatakan positif dengan gerakan LGBT dan 27 menyatakan sikap negative.

Abstract

Aim of this study is to describe the frequency of tweets on Twitter opinion related to the pros and cons of the LGBT movement. The core method research on text mining and sentiment analysis wordclouds with R was applied for this research. Bags of word method is done to assemble the sentiment of twitter. Negative and positive word of pure change of sentiment lexicon Bing Liu are adapted to Indonesian. Stop words that are part of the analysis of text mining is also made to meet up the needs of analysis. The results of the LGBT movement is quite consistent with the current conditions, the results of wordcloud symmetrical with the analysis of sentiment, where the frequency of words such as "cohabiting", "moral", "Indonesia", "conversation", "jail", "move", "challenged", "borderline" is high, and as many as 379 is neutral tweet opinion, 79 expressed positive with the LGBT movement and 27 expressed negative attitudes toward it.

© 2017 Universitas Negeri Semarang

✉Alamat korespondensi:
Fakultas Psikologi, Universitas Widya Dharma, Klaten
hartantopaud@unwidha.ac.id

PENDAHULUAN

Pemberian nama abad 21 sebagai abad informasi modern tidaklah salah. Sebagai contoh penggunaan media sosial seperti Facebook, Twitter, Pinterest, dan Youtube pada era abad informasi ini mengalami kenaikan yang sangat drastis dibandingkan dengan tahun – tahun sebelumnya. Data dari tahun 2015 saja memaparkan pengguna Twitter ada lebih dari 285 juta (Alim, 2015). Mayoritas pengguna dari remaja berumur 18 sampai orang dewasa umur 60 tahun. Penggunaan Twitter dapat mencapai 400 (tweets) lebih dalam seharinya. Twitter juga digunakan dalam ranah sosial-politik seperti dalam pergerakan sosial bersama, melaporkan informasi seperti kemacetan, kondisi cuaca, bencana alam dan memberikan peringatan terhadap suatu kejadian yang akan dihadapi. Beberapa jurnal juga meneliti mengenai keutamaan Twitter dalam melakukan klasifikasi dan analisis dalam bidang pendidikan, khususnya terhadap perilaku bullying (Hon & Varathan, 2015; Ming-Xu, Zhu & Bellmore, 2012), meskipun begitu sebuah analisis teks tidak harus dengan media sosial tertentu, pemilihan pada media sosial Twitter, karena kelebihan Twitter memakai karakter dibawah 140, format pesan yang informal, teks pesan yang sangat banyak setiap harinya, berasal dari individu dengan latar belakang yang sangat bervariasi, dan mudah digunakan (Nahar, Li, & Pang, 2013).

Microblogging seperti Twitter digunakan oleh semua orang untuk melakukan penilaian dan mengeluarkan opini mengenai segala sesuatu, melakukan posting dan melakukan rating dengan opini yang berbeda – beda. Misalnya penggunaan Twitter dalam ranah ekonomi juga terkait dalam melihat perkembangan/trend pasar dan marketing. Melihat opini publik melalui Twitter mengenai suatu jasa atau barang dan membuat rating sehingga perusahaan yang bersangkutan dapat melakukan review dan perbaikan dalam

produk/jasanya (Pak & Paroubek, 2010). Perubahan penggunaan internet dalam hal telekomunikasi mengalami perubahan apabila dulu terfokus pada cara “tradisional” seperti email/surel dan blog maka sudah berubah ke microblogging view karena format pesan yang santai, responsive dan fleksible. Pemakaian Twitter mayoritas mengenai opini atau pendapat pribadi sampai pernyataan – pernyataan yang lebih umum klarifikasi terhadap sesuatu (public statement).

Berdasarkan latar belakang fungsi utama Twitter, maka penelitian ini berfokus pada berita yang memiliki tweet melonjak drastis. Gerakan LGBT yang muncul di penghujung tahun 2016 dan masih hangat sampai penelitian ini ditulis. Perilaku dan gerakan LGBT sukar untuk dikategorikan sebagai fenomena yang benar – benar mengalami masa puncak dan mengalami masa turun. Hal tersebut disebabkan oleh tipikal gerakannya yang murni bersifat latent dan asimetris. Sehingga sangat susah ditebak dan dideskripsikan memakai metode yang biasa. Di sisi lain, menjadi tujuan dari penelitian ini adalah analisis Twitter dengan *text mining* menitik fokuskan pada pendekatan yang sama sekali berbeda, dengan begitu pendekatan yang berbeda kemungkinan menghasilkan jenis hasil yang berbeda pula. Penelitian serupa melakukan *text mining* pada Twitter juga pernah dilakukan oleh Sarwani dan Mahmudi (2015) namun dengan piranti lunak dan metode berbeda (*naïve bayes classifier*).

Twitter berdiri pada bulan Maret 2006 oleh Jack Dorsey, sedangkan prototype perdana situs daringnya diluncurkan pada bulan Juli. Di Twitter, pengguna tak terdaftar hanya bisa membaca kicauan, sedangkan pengguna terdaftar bisa menulis kicauan melalui GUI situs web, pesan singkat (SMS), atau melalui berbagai aplikasi dari perangkat seluler.

Eskalasi Twitter pesat dan dengan cepat dapat meraih popularitas di seluruh dunia. Hingga bulan Januari 2013, nyatanya

terdapat lebih dari 500 juta pengguna terdaftar di Twitter, 200 juta di antaranya adalah pengguna aktif. Karakteristik penggunaan Twitter mengalami kenaikan umumnya pada saat terjadinya peristiwa-peristiwa populer. Hal ini menyebabkan posisi Twitter naik ke peringkat kedua sebagai situs jejaring sosial yang paling sering dikunjungi di dunia, dari yang sebelumnya menempati peringkat dua puluh dua. Semakin naiknya popularitas Twitter menyebabkan layanan ini telah dimanfaatkan untuk berbagai keperluan dalam berbagai ranah, misalnya sebagai sarana protes, kampanye politik, sarana pembelajaran, dan sebagai media komunikasi darurat. Twitter juga dihadapkan pada berbagai masalah dan kontroversi seperti masalah keamanan, pendidikan (*bullying*) dan privasi pengguna (Ohri, 2014).

Text mining (menambang teks) merupakan analisis teks dimana sumber data biasanya di dapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan, keterkaitan dan kelas antar dokumen (Leismester, 2015). Definisi lain, *text mining* melingkupi sebuah proses ekstraksi informasi yang terpola yang berasal dari sejumlah besar sumber data teks, seperti dokumen Word, PDF, kutipan teks, atau bahkan sms (tweet). Teks mining dibagi menjadi 2 tahap, pertama diawali dengan merubah data teks unstruktur ke data semi atau terstruktur dilanjutkan dengan melakukan ekstraksi informasi yang diteliti dari data teks terstruktur. *Text mining* aplikasinya sangat luas, karena kefleksibelan metode ini, bisa digunakan untuk melakukan analisis verbatim misalnya.

Text mining bisa dilakukan klasifikasi (classifier) maupun hanya dengan melihat frekuensi (*wordcloud*) dan dilanjutkan dengan melakukan sentimen analisis (Pathak, 2014). Penelitian ini menggunakan analisis *text mining* pada data unstructured dengan bantuan

open-source R. Dalam melakukan analisis teks, R merangkai data yang acak sehingga menjadi data semi-struktur atau terstruktur dan data siap di lakukan transformasi. Transformasi teks (beberapa menyebut *pre-processing texts*) itu mengubah teks asli kedalam bentuk teks yang lebih sederhana dan mudah dianalisis. Teks mining untuk data web harus menggunakan API dengan web yang akan ditambang.

Beberapa kendala dalam melakukan *text mining* adalah dalam pemakaian bahasa. Software analisis text sejauh ini hanya dapat digunakan secara optimal dalam bahasa Inggris. Hal itu juga yang mempengaruhi penelitian – penelitian mengenai *text mining* berbahasa Indonesia sangat jauh lebih sedikit dibandingkan dengan penelitian dengan metode yang lain. R memiliki fleksibilitas dalam melakukan data mining twitter dengan melakukan settingan hanya untuk yang menggunakan bahasa Indonesia.

Opini, sentiment atau kata yang sering didengar sentimental adalah kata/trait yang melekat pada variable seperti perilaku, evaluasi, appraisal, mood, afeksi, subjective belief/belief system dan feeling. Variabel diatas (dan belum semua) nyatanya sangat mempengaruhi bagaimana kita dalam melihat suatu peristiwa dan mengambil keputusan atau opini. Kemampuan individu dalam budaya kolektif terkadang melihat keputusan atau opini orang lain dulu sebelum memberikan opini pribadi mengenai suatu kejadian. Dalam sosial media data yang ditampung mengenai opini berkembang sangat pesat, karena opini lebih cepat disalurkan dan mudah melakukan akses ke opini orang lain oleh karena itu analisis sentiment sangat diperlukan (Liu, 2015).

Analisis opini atau sentiment merupakan bagian riset dari *text mining*. Sentimen analisis merebak sekitar tahun 2003 (Pang & Lee, 2008), setelah itu banyak penelitian yang mencakup area ini. Aplikasi hasil penelitian ini sangat berguna dalam

banyak bidang, namun ada beberapa bidang mayor yang paling terbantu, yaitu dalam dunia web-online (mengenai pelayan, maupun marketing), periklanan, bisnis dan keamanan negara (intelligence). Beberapa perusahaan profit juga mengembangkan tool sendiri untuk memberikan jasa analisis sentiment (Twitratr, tweetfeel dan Social Mention). Analisis sentiment merupakan tipologi area dari ilmu Natural Language Processing yang bergerak secara kontinum mulai dari level/tahap klasifikasi teks, sampai mereview polaritasnya (Kouloumpis, Wilson, & Moore, 2011). Riset yang paling sering dilakukan terhadap analisis Twitter adalah pada tahap klasifikasi.

LGBT merujuk pada suatu gerakan mengenai usaha pelegalan genderisasi. Dimana kaum dari gerakan LGBT adalah melakukan support dan dukungan bagi setiap manusia yang menginginkan pernikahan sejenis. Penggunaan nama atau istilah LGBT juga sebenarnya masih ambigu karena menuai pro kontra didalam anggotanya dan bahkan secara umum diluar komunitasnya. Seperti dilansir Wikipedia bahwa ada sebagian kaum homoseksual dan lesbian menolak individu dengan transgender dan transseksual dianggap sebagai satu komunitas dengan mereka, bagi kaum homoseks dan lesbi meeka adalah tidaklah sama. Walaupun begitu penggunaan kata LGBT sudah melekat kuat dalam masyarakat umum dimana melukiskan mengenai suatu komunitas individu yang menyukai sesama jenis dan menginginkan pelegalan kepada Negara dan Bangsa mengenai hubungannya tersebut. Pantauan peneliti mengenai topic #LGBT, #homoseks, dll terlihat cukup deras opininya dalam Twitter. Terlepas dari baik buruknya gerakan LGBT, kewenangan itu terlepas dari tujuan peneliti yang berusaha menggambarkan secara faktual opini masyarakat umum mengenai gerakan LGBT. Meskipun begitu perilaku dan gerakan LGBT adalah perilaku yang secara psikologis dapat dikatakan menyimpang. Bagaimanapun juga LGBT, kaum homoseks

dan lesbian tidak bisa jika diperlakukan seperti kebebasan “pilihan” setiap orang untuk berpartisipasi atau tidak, karena sifatnya yang “menular” maka dorongan untuk melakukan hal tersebut bisa disamakan dengan penggunaan narkoba dan sejenisnya (apabila dianggap pilihan maka seorang individu memakai narkoba dia tidak merugikan orang lain, sama seperti ketika orang berkomitment jadi homo/lesbi). Namun nyatanya Negara melarang narkoba sampai sedemikian ketatnya maka seyogyanya perilaku menyukai sesama jenis juga bisa diperlakukan sebagaimana mestinya.

Gerakan LGBT subur makmur melalui media social, dimana di dalam dunia maya tersebut LGBT bisa dengan leluasa melakukan misi dan visinya. Esensi pesan mengenai apa itu pilihan hidup LGBT dapat tersampaikan, namun tanpa melibatkan eksistensi aslinya. Sejarah gerakan LGBT sudah tampak sejak lama, tidak bisa dipastikan sejak kapan muncul namun beberapa teori menduga ikut berkembang seiring perkembangan dan evolusi manusia. Kaum LGBT masuk ke dalam pikiran remaja lewat film dan budaya yang seolah mengijinkan bahwa LGBT itu tidaklah salah dan hal yang lumrah. Jelas, dampak dari penggunaan internet juga sangat besar dalam menyumbang LGBT. Perbedaan mendasar adalah apabila dulu kaum LGBT tertutup dan meyakini bahwa dirinya salah dan berusaha sembuh, sekarang dengan adanya propaganda dan gerakan – gerakan yang massif mereka malah berani tampil dan semakin percaya diri. Salah satu efek negatif yang perlu diwaspadai dari media sosial untuk mengkampanyekan sebuah gerakan tertentu kemungkinan adalah efek *subliminal messages*. SM begitu disingkat, adalah pemasukan pesan – pesan secara tersembunyi yang diterima individu secara tidak sadar dan langsung masuk menuju alam bawah sadar. Namun konten dari pesan ini exist dan ada didalam alam bawah

sadar, dan secara tidak langsung mempengaruhi cara berpikir dan berperilaku.

METODE

Penelitian ini termasuk kedalam *machine learning* analisis. Berbeda dengan tipologi penelitian pada umumnya, *machine learning* lebih kearah klasifikasi, clustering dan kalkulasi prediktif terhadap data (*data mining*) ataupun *big data*. Dewasa ini *data mining* kian pesat penggunaannya di seluruh penjuru dunia. *Data mining* (KDD) merujuk pada tehnik analisis data dari berbagai sumber, sekilas data mining ada kesamaan dengan statistik dan ada pula perbedaannya. Kesamaan dengan ilmu statistik adalah memiliki tujuan untuk menyimpulkan hasil dari data yang diakuisisi, perbedaannya muncul dari jenis/ukuran sumber data dan hipotesis. Sedangkan *big data* mengacu pada jenis data yang memiliki karakteristik tertentu (4V). *Big data* merupakan himpunan data yang sangat besar dan kompleks yang menjadikannya sulit ditangani jika memakai manajemen data biasa atau tradisional.

Metode penelitian ini menggunakan *scraping* data dari Twitter kemudian dianalisis dengan metode *text mining*, sebelumnya dengan menggunakan API antara R dan Twitter. Metode *text mining* memerlukan banyak modifikasi apabila diadopsikan dalam banyak budaya dan bahasa. *Text mining* dalam tweet bahasa Indonesia jelas sangat berbeda dengan *text mining* dalam bahasa Inggris atau German, dan sayangnya metode *text mining* dalam bahasa Indonesia belum ada/belum pernah dilakukan. Perbedaan ini nantinya juga merubah beberapa kode dalam perintah R. Penelitian ini terbagi menjadi 2 tahap, tahap yang pertama adalah melakukan frekuensi analisis *wordcloud* dan tahap yang kedua dilanjutkan dengan sentiment analisis.

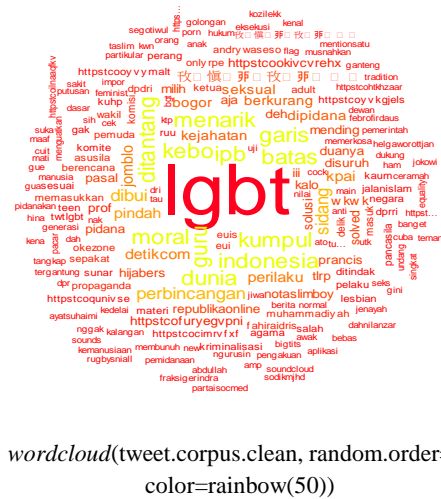
Subjek penelitian ini adalah individu yang aktif menggunakan Twitter dalam menyatakan opini dan pendapatnya terhadap gerakan LGBT. Subjek berkisar dari remaja

sampai orang dewasa pengguna Twitter yang memakai bahasa Indonesia. Data *scraping* dari Twitter hanya membutuhkan posting chat satu subjek dan dikelompokkan pada corpus untuk dilakukan analisis. Pembatasan jumlah chat/tweet subjek dilakukan dengan asumsi parameter statistik dan melalui perintah dalam syntax R. berdasarkan asumsi tersebut jumlah chat/tweet berkisar 300 - 500 sudah sangat mencukupi untuk melakukan analisis teks dan *wordcloud*. Penelitian sejenis banyak dilakukan dengan beberapa software yang berbeda pula, seperti penelitian yang dilakukan oleh Talbot, Acheampong dan Wicentowski (2015) yang menggunakan SemEval-2015, Phyton (Carchiolo, Longheu, & Malgeri, 2015), RapidMiner (Blatnik, Jarm, & Meza, 2014) dan WEKA (Witten, Frank, & Hall, 2011). Penelitian ini memakai R sebagai alat statistik utama, karena selain open-source R dikenal karena reliabilitas, fleksibilitas dan output grafis yang sangat melimpah. Untuk bisa melakukan analisis *text mining* R membutuhkan beberapa bantuan packages.

HASIL DAN PEMBAHASAN

Peneitian dilakukan dengan terlebih dahulu melakukan registrasi kepada Twitter dan memperoleh kode TwitterAPI. Setelah itu dilakukan *synchronize* GUI antar muka dengan *open-source* R. Perbedaan bahasa juga menjadi halangan dalam *text mining* dan sentiment analysis. Hampir mayoritas software di design untuk bahasa Inggris, sehingga membutuhkan modifikasi perintah supaya bekerja dalam bahasa Indonesia. Peneliti mencari solusi atas permasalahan tersebut, dengan inisatif mengganti kode penggunaan bahasa pada CLI R. Penggunaan bahasa Indonesia mengikuti kaedah ISO-639 yaitu dari Inggris = "en", berubah menjadi Indonesia = "id". Ternyata dengan melakukan perubahan kode tersebut semua perintah dalam analisis Twitter akan dialihkan pada tweet yang khusus memakai bahasa Indonesia. Selain itu persiapan lainnya adalah pembuatan

Sentiment Lexicon yaitu daftar kata (*words list*) yang dibuat secara manual—sesuai dengan konten bahasa—berfungsi sebagai pembandingan dari text yang sedang dianalisis setelah dibandingkan itu menjadi alat untuk melakukan analisis text. *Sentiment Lexicon* dibukukan oleh Bing Liu yang berisi sekitar 2000 kata positive dan 4000 jadi total sekitar 6800 kata (Liu, 2015).



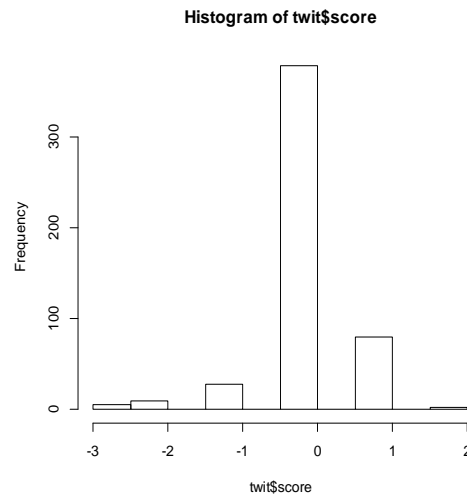
Gambar 1. Word Cloud LGBT

Wordcloud berhasil dilakukan dan dapat dianalisis kata – kata seperti “kumpul kebo”, “moral”, “Indonesia”, “perbincangan”, “dibui”, “pindah”, “ditantang”, “garis batas”, menjadi kata yang frekuensinya tinggi dalam perbincangan mengenai tumbuh kembangnya gerakan LGBT di Indonesia. Sedangkan kata seperti “kpai”, “sidang”, “prancis”, “perilaku” menjadi kata dengan frekuensi kedua yang muncul dalam analisis. Analisis diatas masih memiliki kekurangan dalam tampilan grafik *wordcloud*. Pertama dari tata bahasa masih terdapat beberapa stopword dalam bahasa Indonesia, dimana dalam syntax R memang tidak memberikan untuk bahasa Indonesia, walaupun sudah ada beberapa dari bahasa selain Inggris. Kumpulan stopwords peneliti rangkum dari kamus besar bahasa Indonesia dan tambahkan perintah dari beberapa diskusi di situs www.stackoverflow.com sehingga bisa

fit dengan konsep *text mining* yang peneliti inginkan.

Selanjutnya adalah tahap kedua yaitu sentiment analysis, dengan berdasar pada data diatas. *Sentiment Lexicon* yang digunakan adalah gabungan dari kata – kata positive dan negative untuk gerakan LGBT dan opini masdevid (terjemahan Bing Liu Sentiment Lexicon) yang diunduh dari gihub.com (<https://github.com/masdevid/ID-Opinion> Words). Meskipun begitu positive word dan negative word harus dirubah sedemikian ketat sampai mendekati pokok bahasan dari topic twitter. Analisis sentiment dalam R bisa dengan banyak tehnik dan cara, pada penelitian ini peneliti menggunakan pendekatan dari Jeffrey Breen (github.com).

```
> table(twit$score)
-3 -2 -1 0 1 2
  5  9 27 378 79 2
> mean(twit$score)
[1] 0.046
> median(twit$score)
[1] 0
```



```
> hist(twit$score)
```

Hasil yang diperoleh mayoritas masuk dalam ranah netral, plus 1, dan minus 1. Sentiment lexicon dari bahasa Indonesia bekerja dengan baik, namun ada yang ditaksir oleh R sebagai text yang netral dibuktikan

dengan bobot 0 (>300). Namun meskipun begitu terlihat 79 tweet menyatakan positif, artinya analisis diatas ditemukan tweet positif mengenai perilaku LGBT yang disinyalir adanya dukungan dari perilaku tersebut, dengan kata lain perilaku LGBT ternyata juga memiliki dukungan di Negara Indonesia. Grishman (2012), secara teknis, menambahkan metode *bags of word* memiliki kelemahan yaitu *list words* akan berbeda untuk topik yang lain. Lebih detail lagi *bag of word* mengesampingkan masalah semantik dan sinonim kata, memproduksi hanya nomor kata. Terbatasnya *bag of word* di atasi oleh *word2vec* yang memiliki kemampuan tinggi menebak persamaan kata. Selebihnya data menyatakan tidak suka sampai bobot skor -3, dimana skor suka maksimal 2. Sejauh ini kemampuan R dalam melakukan sentiment analysis dianggap adekuat dan cukup powerful bersanding dengan statistikal packages komersial seperti Lexalytics dan Clarabridge (berkisar \$2499)

SIMPULAN

Paling tidak ada 2 cara dalam melakukan sentiment analisis, yaitu pertama dengan metode “*bag of words*” dimana memakai bantuan algoritma khusus dan words list yang berperan sebagai “filter” dari teks yang dianalisis (polarity) dan melakukan ekstraksi sentiment atau opini sebagai hasilnya. Kedua dengan metode statistik konsep hubungan dimana mengukur polaritas bukan dari word list namun dengan cara melakukan “comparing” terhadap satu set variable baru setelah itu dilakukan ekstraksi opini/sentiment.

Cara kedua biasa digunakan oleh perusahaan besar dimana sentiment analisis dilakukan dalam jumlah besar (big data) sehingga memungkinkan untuk meneruskan analisis lain terhadap suatu variabel. Saran pertama untuk penelitian mengenai konsep ini lebih tepatnya mengenai pembukuan sentiment lexicon dan stopwords dalam

bahasa Indonesia yang lebih detail. Pemilihan metode yang berbeda mungkin juga menyuguhkan hasil yang beragam seperti MaxEn, SVM dan Naïve Bayes. Saran kedua adalah pemusatan fokus pada ranah – ranah psikologi, dimana hasil penelitian ini bisa menjadi data rujukan dari langkah riset yang lebih besar.

DAFTAR PUSTAKA

Alim, S. (2015). Analysis of Tweets Related to Cyberbullying: Exploring Information Diffusion and Advice Available for Cyberbullying Victims. *International Journal of Cyber Behavior, Psychology and Learning*, 5(4), 31-52, October-December 2015.

Blatnik, A. jarm, K. & Meza, M. (2015). Movie sentiment analysis based on public tweets. *ELEKTROTEHNI`SKI VESTNIK* 81(4) : 160–166, 2014.

Carchiolo, V. Longheu, A. & Malgeri, M. (2015). *Using Twitter Data and Sentiment Analysis to Study Diseases Dynamics*. M.E. Renda et al. (Eds.): ITBAM 2015, LNCS 9267, pp. 16–24, 2015.

Hon, L, C. & Varathan, K, D. (2015). Cyberbullying Detection System on Twitter. *International Journal of Information System and Engineering*. Vol 1 (No.1), April 2015.

Kouloumpis, E. Wilson, T. & Moore, J. (2011). Twitter Sentiment Analysis – The Good and The Bad and The OMG!. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*.

Liu, B. (2015). *Sentiment Analysis*. Cambridge University Press. www.cambridge.org

Leismester, C. (2015). *Mastering Machine Learning with R*. Published by Packt Publishing Ltd. Livery Place 35 Livery Street

Ming-Xu, J. Zhu, X. & Bellmore, A. (2012). *Fast Learning for Sentiment Analysis on Bullying*. WISDOM '12, August 12 2012,

Beijing, China. Copyright 2012 ACM 978-1-4503-1543-2/12/08

Ohri, A. (2014). *R for Cloud Computing*. Springer Science+Business Media New York 2014

Pak, A. & Paroubek, P. (2010). Twitter as Corpus for Sentiment Analysis and Opinion Mining. In *Proceeding in LREC*

Pang, B. & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends In Information Retrieval* Vol. 2, Nos. 1–2 (2008) 1–135 C 2008 B. Pang and L. Lee

Pathak, M, A. (2014). *Beginning Data Science with R*. Springer International Publishing Switzerland 2014.

Talbot, R. Acheampong, C. & Wicentowski, R. (2015). SWASH: A Naive Bayes Classifier for Tweet Sentiment Identification. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 626–630.

Witten, I.H. Frank, E. & Hall, M, A. (2011). *Data Machine – practical machine learning tools and technique*. Morgan Kaufmann Publisher is in imprint Elsevier. USA

<https://github.com/masdevid/ID-OpinionWords> Grishman, R. 2012.cs.nyu.edu/~fergus/teaching/vision_2012/9_BoW.pd