Deteksi *Bot Spammer* pada *Twitter* Berbasis *Sentiment Analysis* dan *Time Interval Entropy*

Christian Sri Kusuma Aditya[1], Mamluatul Hani'ah[2], Alif Akbar Fitrawan[3], Agus Zainal Arifin[4], Diana Purwitasari[5]
Jurusan Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember
Jl. Teknik Kimia, Gedung Teknik Informatika, Kampus ITS Sukolilo, Surabaya 60111, Jawa Timur
E-mail: [1]christian.s.k.aditya@gmail.com, [2]mamluatul14@mhs.if.its.ac.id, [3]alif14@mhs.if.its.ac.id, [4]agusza@cs.its.ac.id, [5]diana@if.its.ac.id

Masuk: 22 Desember 2015; Direvisi: 29 Januari 2016; Diterima: 3 Februari 2016

Abstract. *Spam is an abuse of messaging undesired by recipients. Those who send spam are called spammers. Popularity of Twitter has attracted spammers to use it as a means to disseminate spam messages. The spams are characterized by a neutral emotional sentiment or no particular users' preference perspective. In addition, the regularity of tweeting behavior periodically shows automation performed by bot. This study proposes a new method to differentiate between bot spammer and legitimate user accounts by integrating the sentiment analysis (SA) based on emotions and time interval entropy (TIE). The combination of knowledge-based and machine learning-based were used to classify tweets with positive, negative and neutral sentiments. Furthermore, the collection of timestamp is used to calculate the time interval entropy of each account. The results show that the precision and recall of the proposed method reach up to 83% and 91%. This proves that the merging SA and TIE can optimize overall system performance in detecting Bot Spammer.*
Keywords: *bot spammer, twitter, sentiment analysis, polarity, entropy*

Abstrak. *Spam merupakan penyalahgunaan pengiriman pesan tanpa dikehendaki oleh penerimanya, orang yang mengirimkan spam disebut spammer. Ketenaran Twitter mengundang spammer untuk menggunakannya sebagai sarana menyebarluaskan pesan spam. Karakteristik dari tweet yang dikategorikan spam memiliki sentimen emosi netral atau tidak ada preferensi tertentu terhadap suatu perspektif dari user yang memposting tweet. Selain itu keteraturan waktu perilaku saat memposting tweet secara periodik menunjukkan otomatisasi yang dilakukan bot. Pada penelitian ini diusulkan metode baru untuk mendeteksi antara bot spammer dan legitimate user dengan mengintegrasikan sentimen analysis berdasarkan emosi dan time interval entropy. Pendekatan gabungan knowledge-based dan machine learning-based digunakan untuk mengklasifikasi tweet yang memiliki sentimen positif, negatif dan tweet netral. Selanjutnya kumpulan timestamp digunakan untuk menghitung time interval entropy dari tiap akun. Hasil percobaan menunjukan bahwa precision dan recall dari metode yang diusulkan mencapai 83% dan 91%. Hal ini membuktikan penggabungan Sentiment Analysis (SA) dan Time Interval Entropy (TIE) dapat mengoptimalkan performa sistem secara keseluruhan dalam mendeteksi Bot Spammer.*
Kata Kunci: *bot spammer, twitter, sentiment analysis, polarity, entropy*

1. Pendahuluan

Situs *microblogging* telah menjadi alat komunikasi yang sangat populer di kalangan pengguna *internet*. Setiap hari jutaan pesan muncul di situs *web* populer yang menyediakan layanan *microblogging* seperti *Twitter*, *Tumblr*, dan *Facebook*. *Twitter* adalah sebuah situs *web* yang dimiliki dan dioperasikan oleh *Twitter Inc.*, menawarkan jaringan sosial berupa *microblogging* sehingga memungkinkan penggunanya untuk mengirim dan membaca pesan di *Twitter*. Tidak seperti *Facebook*, *LinkedIn*, dan *MySpace*, *Twitter* merupakan sebuah jejaring sosial yang dapat digambarkan sebagai sebuah *graph* berarah yang berarti bahwa pengguna

dapat mengikuti pengguna lain, namun pengguna kedua tidak diperlukan untuk mengikutinya kembali. *Twitter* mengizinkan penggunanya untuk membaca dan menulis pesan singkat yang dibatasi maksimal 140 karakter. Pesan singkat yang sering disebut *tweet* tersebut kebanyakan bersifat publik dan dapat dilihat oleh pengguna lain. Menurut penelitian Takhteyev, dkk. (2012) hanya terdapat sekitar 10% dari pengguna *Twitter* yang memproteksi *tweet* mereka.

Twitter dengan pengguna lebih dari 500 juta dan 400 juta *tweet* perhari memungkinkan pengguna untuk berbagi pesan menggunakan *tweet*. Pengguna *Twitter* menulis tentang kehidupan mereka, berbagi opini tentang berbagai topik dan membahas isu-isu yang terjadi pada saat ini. Aksesibilitas dari berbagai *platform* yang mudah, pengguna *internet* cenderung untuk beralih dari *blog* atau milis ke layanan *microblogging*. Hal tersebut menyebabkan semakin banyak pengguna *Twitter* yang melakukan posting tentang suatu produk dan layanan yang mereka gunakan untuk mengekspresikan pandangan mereka. *Twitter* dapat menjadi sumber data pendapat dan sentimen masyarakat. Seiring dengan kepopuleran dan potensi dari *Twitter* di dunia *internet* menyebabkan para *spammer* mulai melirik untuk membanjiri pesan *spam* demi keuntungan pribadinya. Aktifitas *spam* dilakukan atas berbagai tujuan. Salah satunya adalah tujuan komersial berupa iklan dan promosi. *Spam* ini tidak memerlukan *mailing list* untuk mencapai para pelanggan-pelanggan yang diinginkan, oleh karena itu *spam* dikirimkan dengan biaya operasi yang sangat rendah. *Spam* bisa berisi pesan singkat atau menanam *link* saja yang sebagian besar tidak berfokus kepada materi *posting* yang sedang disajikan, namun jika dilakukan secara terus-menerus akan mengganggu pengguna yang menerima ataupun sekedar melintas. Dari aspek teknis, pesan yang dikirim sekaligus dan terus-menerus akan membutuhkan sumber daya *(resource)* yang sangat banyak, baik dari segi kemampuan *server* maupun *bandwidth*, hal ini dapat mengakibatkan beban *server* yang tinggi yang dapat mengakibatkan *server down*. Tak jarang pula pelaku menggunakan robot ataupun aplikasi yang secara otomatis mengirim pesan *spam.*

Pengguna *Twitter* diklasifikasikan menjadi beberapa kategori yaitu manusia *(legitimate user)*, *cyborg*, dan *bot*. Fitur-fitur *Twitter* seperti waktu pengguna mengunggah *tweet*, *tweet content*, dan *account properties* adalah fitur yang dapat digunakan untuk mengidentifikasi manusia, *cyborg*, atau *bot*. Diantara fitur-fitur tersebut, waktu pengguna untuk mengunggah dapat digunakan untuk mencari *interval entropy*. *Interval entropy* ini dapat menghasilkan tingkat akurasi tinggi jika dibandingkan fitur *tweet content* dan *account properties*.

Program otomatisasi atau yang dikenal sebagai *bot* merupakan kependekan dari nama robot. *Bot* tidak membutuhkan campur tangan manusia dalam melakukan pekerjaan rutinitasnya setiap waktu. *Bot spammer* secara otomatis menghasilkan *spam* pada interval waktu tertentu menggunakan *scheduler* pekerjaan. Pengeloaan akun *spam* secara manual dapat menyebabkan biaya yang tinggi sehingga dengan digunakannya *bot* maka dapat mengurangi biaya. Dengan demikian *spammer* akan lebih mudah untuk menghasilkan pesan *spam* dalam jumlah yang banyak di *Twitter*. Selain untuk menghasilkan pesan dalam jumlah yang banyak beberapa *bot* diciptakan untuk menyebarkan pesan yang berbahaya (Heron, 2009).

Twitter sendiri memiliki mekanisme untuk penanganan *spam* dengan mengajak pengguna *Twitter* untuk melaporkan pesan *spam* dan akun yang terindikasi sebagai *bot*. Namun cara ini dinilai memiliki kelemahan apabila laporan pengguna *Twitter* yang dikumpulkan ternyata laporan palsu, kesalahan pelabelan akun *legitimate* sebagai *spammer* atau *bot* dapat menurunkan kredibilitas *Twitter*. Beberapa penelitian telah dilakukan mengenai otomatisasi (*bot*) dan deteksi *spam*, untuk membantu mengurangi munculnya *spam* khususnya di *Twitter*.

Penelitian yang dilakukan oleh Chu, dkk. (2012) dapat mengidentifikasi akun manusia, *cyborg*, atau *bot* dengan cara mengamati perbedaan kebiasan perilaku menge*tweet*, konten dari *tweet*, serta karakteristik akun seperti jumlah *follower*, *following*, dan *retweet*. Pada konten dari *tweet* dapat dilakukan deteksi emosi/sentimen emosi dengan menggunakan *machine learning* atau polaritas. Penelitian tentang polaritas (Lima, dkk., 2015) dan sentimen emosi (Mohammad, dkk., 2014) pada *Twitter* juga telah banyak dilakukan. Penelitian Lima, dkk. (2015) memperkenalkan *framework* PAFRA untuk mengklasifikasikan polaritas *tweet* terhadap suatu topik.

Jika dilihat dari karakteristiknya, beberapa *tweet* yang masuk kategori *spam* sering kali di-*posting* secara otomatis dimana *tweet* diunggah secara teratur dalam waktu yang dekat. Selain itu *tweet* tersebut seringkali tidak memiliki ungkapan ekspresi, sedangkan *legitimate user* cenderung mem-*posting tweet* yang memiliki preferensi tertentu terhadap suatu perspektif dan ideologi yang memiliki ungkapan ekspresi pandangan ataupun opini. Sehingga untuk mendeteksi *bot spammer* pada *twitter* tidak cukup hanya dengan menggunakan *time interval entropy* untuk mendeteksi regularitas kebiasaan *posting tweet*, akan tetapi dibutuhkan metode yang dapat mendeteksi sebuah ungkapan ekspresi pada *tweet*.

Oleh karena itu, pada penelitian ini diusulkan metode baru untuk mendeteksi antara *bot spammer* dan *legitimate user* dengan mengintegrasikan *Sentiment Analysis* (SA) berdasarkan emosi dan *Time Interval Entropy* (TIE). S*entiment analysis* (SA) digunakan untuk mendeteksi ungkapan ekspresi ataupun opini yang terkandung dalam *tweet. Sentiment analysis* (SA) pada penelitian ini menggunakan penggabungan metode *knowledge-based* dan *machine learning-based* untuk mendapatkan *tweet* netral atau yang tidak memiliki sentimen emosi dimana sering muncul pada *tweet* bertipe *spam*. Sedangkan TIE digunakan untuk menangkap keteraturan waktu mem-*posting tweet* yang menunjukkan *tweet* diunggah secara otomatis.

2. Data dan Metode Penelitian

2.1. Data

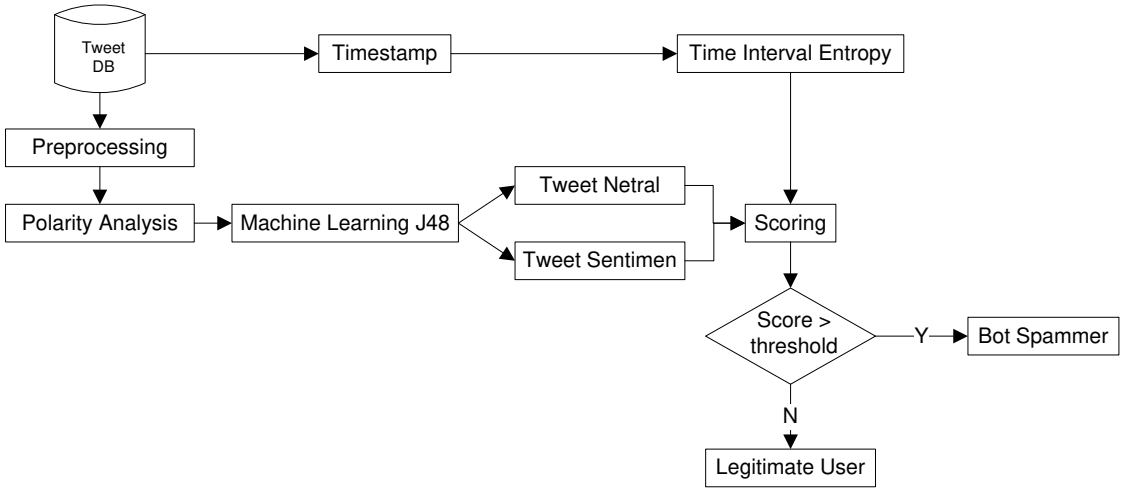
Data *Tweet* dalam penelitian ini diperoleh dengan memanfaatkan API yang disediakan oleh *Twitter*. Dengan memanfaatkan API tersebut dibangun sebuah aplikasi untuk mengambil data *tweet* dari *Twitter* kemudian disimpan ke dalam *database*. Total semua data yang didapat adalah 39 akun pengguna *Twitter* yang ditulis menggunakan Bahasa Indonesia. Terdapat sekitar 500 *tweet* untuk setiap akun. Selanjutnya setiap akun akan diklasifikasikan secara manual antara akun *legitimate user* dan akun *bot spammer.* Dari hasil klasifikasi secara manual didapat sebanyak 14 akun *legitimate user* dan 25 akun *bot spammer*. Jumlah data *training set* yang digunakan sebanyak 11.000 *tweet* dari 21 akun sedangkan untuk *test set* sebanyak 18 akun dengan masing-masing kurang lebih 500 *tweet* tiap akun.

Terdapat beberapa karakteristik yang digunakan untuk melakukan pelabelan manual akun *bot spammer,* karakteristik tersebut adalah sebagai berikut (Yang, dkk., 2011): (1) *Spam* yang berisi *link* aktif. Hal ini di lakukan untuk mempromosikan sebuah *website* dengan cara menautkan *link* aktif berupa URL. (2) *Spam* yang berisi promosi atau menawarkan produk tertentu. *Spam* ini masih berkaitan dengan jenis yang pertama, yaitu berisi *link* aktif yang menawarkan promosi produk tertentu. (3) Kesamaan *tweet* dengan *tweet* sebelumnya. Hal ini dilihat berdasarkan kumpulan *tweet* yang telah di-*posting* oleh pengguna *Twitter*, jika tiap *tweet* memiliki kesamaan konten atau kemiripan kemunculan kata yang digunakan maka akun tersebut dapat dikategorikan sebagai *bot spammer*. (4) Akun baru, *Spammer* biasanya selalu berganti-ganti akun *Twitter*. Usia dari sebuah akun dapat diketahui melalui informasi yang ada pada profil akun yang bersangkutan. (5) *Spam* seringkali memakai banyak *hashtag*. *Hashtag* memudahkan pencarian *tweet*, atau memperbesar peluang untuk menjadi *trending topic* (Verma & Sofat, 2014).

2.2. Metode

Pada penelitian ini diusulkan metode baru untuk membedakan antara *bot spammer* dan *legitimate user* dengan mengintegrasikan *sentiment analysis* (SA) berdasarkan emosi dan *time interval entropy* (TIE). Skema dari usulan kontribusi metode dan proses keseluruhan sistem dapat dilihat pada Gambar 1. Langkah pertama adalah melakukan tahap *preprocessing*. *Preprocessing* terdiri dari empat langkah antara lain *data cleaning*, *tokenizing*, *stopword removal* dan *stemming* yang dilakukan secara berurutan (Miller, dkk., 2014). *Data cleaning* adalah membersihkan *tweet* dari tautan URL, *mention*, *hashtag*, dan simbol RT. *Tokenizing* digunakan untuk mempartisi *tweet* menjadi *token* atau kata. *Stopword removal* dilakukan dengan cara menghapus daftar kata yang dianggap kurang penting, daftar kata *stopword* diambil dari penelitian Tala (2003). Proses terakhir dari *preprocessing* adalah *stemming* yaitu mencari

kata dasar dari tiap kata *tweet* menggunakan algortima yang diusulkan oleh Arifin & Setiono (2002).



Gambar 1. Diagram Blok Metode yang Diusulkan

Setelah melakukan tahap awal pemrosesan teks berikutnya adalah mencari sentimen *tweet* untuk tiap akun. Sebuah sentimen pada dasarnya adalah ungkapan polaritas sebuah teks yang dilabelkan apakah itu berkonotasi positif, netral, dan negatif (Kontopoulos, dkk., 2013). Penentuan sentimen dapat dilakukan pada tingkatan yang berbeda: dokumen (Moraes, dkk., 2013), kalimat (Poria, dkk., 2014) dan kata atau atribut. *Sentiment analysis* atau *opinion mining* adalah studi komputasional dari opini-opini orang, sentimen dan emosi melalui entitas atau atribut yang dimiliki yang diekspresikan dalam bentuk teks (Liu, 2012). Pendekatan SA yang digunakan pada penelitian ini adalah dengan penggabungan metode *Knowledge-based* dan *Machine learning-based.*

Knowledge-based merupakan metode pendekatan SA yang menggunakan bantuan sebuah kamus atau *dictionaries,* seperti *Linguistic Inquiry and Word Count* (LIWC) (Tausczik & Pennebaker, 2010) dan *SentiWordnet* (Esuli & Sebastiani, 2006) yang sering digunakan pada domain teks Bahasa Inggris. Kamus dibentuk oleh kumpulan kata dan diklasifikasikan sesuai nilai polaritasnya. Misalnya, kata bahagia memiliki nilai +1 menunjukkan bahwa kata ini memiliki polaritas positif, atau sebaliknya, kata sedih memiliki nilai -1 menunjukkan polaritas negatif (Montejo-Ráez, dkk., 2014). Bentuk paling sederhana untuk mendapatkan polaritas dari teks adalah dengan menjumlahkan nilai sentimen dari semua kata-kata yang ada dalam teks dan menentukan polaritas yang dihasilkan (Lima, dkk., 2015). Melalui beberapa kali percobaan, ditentukan interval *threshold* untuk kategori netral adalah -1 sampai dengan 1, sedangkan nilai diatas +1 adalah sentimen positif dan nilai dibawah -1 adalah sentimen negatif. Berikut contoh dari *posting tweet* bersentimen netral dan yang bersentimen positif/negatif dapat dilihat pada Tabel 1.

Tabel 1. Polaritas *tweet*

Interval Polaritas	*Tweet*	Sentimen
0	Hai Kak, @alisya_monica, *Follow* @InfoMakassarID yuk.. untuk dapat info paling *update* seputar Makassar.., pasti di *Folback* !	Netral
0	hai kak,@rudhybm, mau dapat info terbaru seputar peluang usaha? *follow* @InfousahaID ya.. Pasti di *Folback* !	
1	hai kak,@fahirafahi,Sudah Terbukti di 17 Negara Cara Belajar Bahasa Asing Tanpa Kursus Info PIN:7D07C8E6 http://t.co/qEc2JHDSYH	
4	RT @tsuroiya: Makin hormat dan kagum sama Bu Susi. Perjalanan hidupnya luar biasa dahsyat, tp cara dia bercerita santai dan rendah hati.	Positif/Negatif
-2	Turut berduka cita atas berpulangnya ananda Maulana (PTIIK-2013). :(Semoga amal ibadahnya diterima olehNya... Aamiin #fb	

Pada penelitian ini daftar kata *lexical database* yang digunakan adalah daftar *lexical* dari penelitian (Hu & Liu, 2004). Daftar tersebut kemudian diterjemahkan dan dilakukan beberapa penyesuaian ke dalam Bahasa Indonesia. Total jumlah keseluruhan *lexical* berisi 3.535 kata dengan daftar jumlah kata sentimen positif sebanyak 1.126 dan jumlah kata sentimen negatif sebanyak 2.409. Contoh daftar kata yang digunakan untuk menentukan polaritas menggunakan pendekatan *knowledge-based* dapat dilihat pada Tabel 2.

Tabel 2. Contoh daftar kata *lexical database*

Kata Positif	Kata Negatif	Emoticon Positif	Emoticon Negatif
adil, arif, alhamdulillah, bahagia, baik, berkah, bersih, bijaksana, cepat, cerdas, damai, empati, fokus, gigih, hebat, inisiatif, jujur, kagum, kuat, lancar, manis, nikmat, pandai, rapi, sahaja, tegas, ulet, wibawa, ...	abai, aneh, apati, bahaya, benci, cemas, cela, dendam, erang, gagal, gila, hina, injak, jahat, keji, korup, lambat, malas, nakal, picik, rakus, sakit, takut, ...	:-), :), =), :D, o:)	:-(, :(, =(, ;(, T_T, -_-

Setelah melalui proses perhitungan interval polaritas, data *tweet* akan digunakan sebagai *training set* untuk proses *machine learning*. *Machine learning-based* adalah metode pendekatan menggunakan algoritma pembelajaran seperti *Naive Bayes* (NB), J48, *Support Vector Machine* (SVM), dan lain sebagainya. Metode ini sering digunakan untuk klasifikasi teks dan dapat menunjukkan efektivitasnya ketika diterapkan pada permasalahan SA. Pendekatan ini memerlukan *training set* untuk pembentukan model klasifikasi dan kemudian dibandingkan dengan data baru yang belum terlabelkan atau disebut *test set* untuk diklasifikasikan (Drucker, dkk., 1999). Hal ini penting dilakukan untuk mengevaluasi kemampuan generalisasi dari algoritma terhadap data baru dan keakuratan mengklasifikasikan. Biasanya, pelabelan *training set* membutuhkan banyak waktu dan tenaga apabila dibuat secara manual, dan sangat tergantung pada persepsi masing-masing individu (Lima, dkk., 2015). Adanya keterbatasan tersebut, digunakan model *hybrid* yaitu menggunakan alat bantu *lexical database* sebagai pelabelan secara otomatis data *tweet* untuk pembentukan *training set* pada *machine learning-based*. Pada penelitian ini digunakan algoritma klasifikasi J48 (Thelwall, dkk., 2010) untuk mendeteksi sentimen setiap *tweet*. Dari setiap dokumen *training* diekstraksi *term*-nya dan diberikan bobot, kemudian *term* tersebut ditetapkan sebagai kata kunci untuk setiap kategori sentimen (Li & Xu, 2014).

Langkah berikutnya, data *timestamp* saat mem-*posting tweet* dari tiap akun *Twitter* dikumpulkan dan dicari interval waktunya. TIE digunakan untuk menangkap pola keteraturan waktu *posting tweet* yang menunjukkan otomatisasi. TIE (H) dihitung dengan menggunakan persamaan (1) dan persamaan (2) (Chu, dkk., 2012). ΔT merepresentasikan interval waktu antar *tweet*, dimana $P\Delta T(\Delta t_i)$ menunjukkan probabilitas interval waktu ΔT_i. Komponen *entropy* dapat mendeteksi waktu periodik yang merupakan indikasi kuat kejadian otomatisasi. Pengguna *twitter* yang memiliki *entropy* lebih rendah dari *threshold* akan diklasifikasikan sebagai *bot spammer* karena nilai entropi rendah dibawah *threshold* menunjukkan perilaku yang teratur. Terakhir, kedua nilai SA dan TIE digabungkan menggunakan persamaan (3) untuk mengklasifikasikan setiap akun pengguna *Twitter* ke dalam kelasnya.

$$H_{\Delta T}(Ti) = -\sum_{i=1}^{nT} P\Delta T(\Delta t_i)\log(P\Delta T(\Delta t_i)) \quad (1)$$

$$P\Delta T(\Delta t_i) = \frac{n\Delta t_i}{\sum_{k=1}^{nT} n\Delta t_k} \quad (2)$$

$$Score_k = \frac{\alpha(1-H_k)+\beta(sa_k)}{\alpha(max(1-H_k))+\beta(max(sa_k))} \quad (3)$$

Untuk setiap akun pengguna *Twitter* k, nilai perhitungan SA dan TIE dikalikan dengan faktor pembobotan untuk mengambil nilai akhir. Variabel α dan β masing-masing menunjukkan faktor pembobotan dari SA dan TIE. Jumlah dari kedua faktor pembobotan variabel bernilai 1 (Perdana, dkk., 2015).

3. Hasil dan Diskusi

Untuk melakukan evaluasi kinerja metode yang diusulkan secara kuantitatif, digunakan perhitungan *precision* dan *recall*. *Precision* adalah jumlah kelompok dokumen relevan dari total jumlah dokumen yang ditemukan oleh sistem. Sedangkan, *recall* diartikan sebagai jumlah dokumen relevan yang ditemukan oleh sistem. *Precision* dan *recall* dapat dilihat pada persamaan (4) dan persamaan (5).

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

Perhitungan *precision* dan *recall* menggunakan kombinasi *True Positive* (TP), *False Positive* (FP), dan *False Negative* (FN). Dalam penelitian ini, TP mengacu pada jumlah akun yang diklasifikasikan dengan benar sebagai *bot spammer*. FP merupakan jumlah akun *legitimate user* yang diklasifikasikan tidak benar sebagai *bot spammer*. Sedangkan FN adalah *bot spammer* yang tidak tepat diklasifikasikan sebagai akun *legitimate user*. Penentuan nilai *threshold* pada tiap skenario uji coba berbeda, hal ini berdasarkan beberapa kali percobaan untuk pencarian hasil *precision* dan *recall* terbaik.

Pada uji coba pertama menggunakan fitur SA, jika *user* memliki nilai sentimen diatas batas *threshold* 0,5, *user* tersebut diklasifikasikan sebagai *bot spammer*. Hasil uji coba pertama dapat dilihat pada Tabel 3. Pada Tabel 3, terdapat beberapa *user* yang tidak tepat terklasifikasikan sesuai *groundtruth*. Pada user nomor 3 yang seharusnya terklasifikasikan sebagai *spam* namun oleh sistem mendapat nilai dibawah *threshold*, hal ini disebabkan karakteristik *user* tersebut cukup mirip dengan *legitimate user* dimana menggunakan beberapa kata yang terindikasi memiliki sentimen. Selain itu beberapa pengguna *Twitter* sering menggunakan singkatan kata, bahasa campuran selain Bahasa Indonesia dan penggunaan kata yang tidak sesuai Ejaan Yang Disempurnakan (EYD), dimana dapat menyulitkan fitur yang diambil serta mengurangi ketepatan klasifikasi. Untuk nilai *precision* dan *recall* penggunaan fitur SA secara berurutan adalah 82% dan 82%.

Untuk uji coba kedua menggunakan fitur TIE, hasil uji coba terlampir pada Tabel 4, terdapat beberapa akun *legitimate user* tidak tepat terklasifikasikan sebagai *bot spammer*, hal ini mengindikasikan bahwa beberapa perilaku *legitimate user* sering memiliki pola keteraturan interval waktu yang sama dengan *bot spammer* dalam memposting *tweet.* Beberapa contoh akun *twitter* yang tidak tepat terklasifikasi dengan menggunakan fitur TIE adalah akun berjenis media berita, dimana akun media berita seringkali mem-*posting tweet* secara teratur sesuai jadwal meskipun tanpa menggunakan robot ataupun aplikasi yang secara otomatis. Dengan demikian, penggunaan fitur TIE juga tidak cukup optimal untuk membedakan *legitimate user* dan *bot spammer* dengan nilai *precision* dan *recall* secara berurutan adalah 75% dan 82%.

Pada uji coba ketiga, dilakukan penggabungan kedua fitur SA dan TIE dengan penggunaan bobot rasio perbandingan α dan β bernilai 1:1 dengan nilai *precision* dan *recall* secara berurutan 83% dan 91%. Berikut terlampir hasil uji coba ketiga pada Tabel 5.

Berdasarkan hasil yang telah dipaparkan metode yang diusulkan memiliki nilai *precision* sebesar 83.00% dan *recall* sebesar 91%. Perbandingan dari hasil uji coba semua metode dapat dilihat pada Gambar 2 dimana terdapat perbandingan nilai *precision* dan *recall*. Dari Gambar 2 dapat dilihat bahwa penggunaan gabungan fitur SA+TIE dapat mengoptimalkan nilai *precision* dan *recall* dibanding kedua uji coba sebelumnya yang mengindikasikan bahwa kedua fitur dapat saling terintegrasi dan menutupi kekurangan dari penggunaan satu fitur, SA atau TIE.

Tabel 3. Hasil klasifikasi sistem menggunakan SA.

User ID	Nilai	*Class*	*Ground truth*
1	0,12	L	*Spam*
2	0,19	L	L
3	0,21	L	*Spam*
4	0,96	S	S
5	0,92	S	S
6	0,87	S	S
7	0,79	S	S
8	0,62	S	L
9	0,91	S	S
10	0,13	L	L
11	0,71	S	S
12	0,98	S	S
13	0,89	S	S
14	0,93	S	S
15	0,37	L	L
16	0,57	S	L
17	0,42	L	S
18	0,39	L	S
Threshold : 0,5			

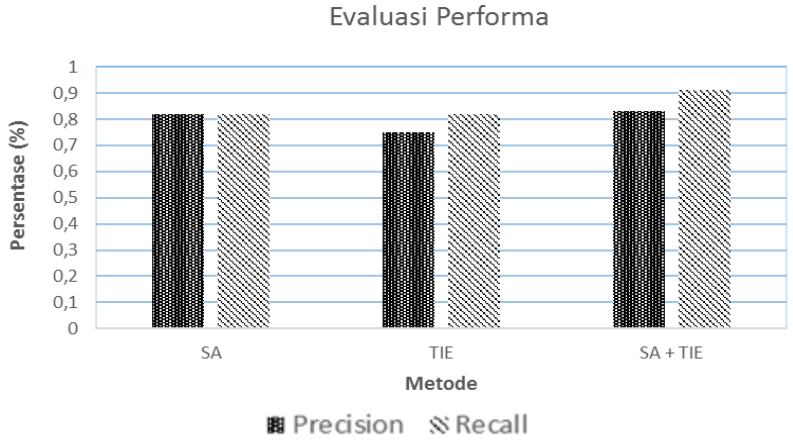
L: *Legitimate*, S: *Spam*

Tabel 4. Hasil klasifikasi sistem menggunakan TIE

User ID	Nilai	*Class*	*Ground truth*
1	0,62	S	S
2	0,71	L	L
3	0,69	L	S
4	0,03	S	S
5	0,02	S	S
6	0,02	S	S
7	0,22	S	S
8	0,04	S	L
9	0,21	S	S
10	0,51	L	L
11	0,02	S	S
12	0,01	S	S
13	0,02	S	S
14	0,02	S	S
15	0,30	S	L
16	0,29	S	L
17	0,02	S	S
18	0,51	L	S
Threshold : 0,3461			

Tabel 5. Hasil klasifikasi sistem menggunakan SA dan TIE

User ID	Nilai	*Class*	*Ground truth*
1	0,26	S	S
2	0,25	L	L
3	0,27	L	S
4	0,99	S	S
5	0,98	S	S
6	0,95	S	S
7	0,81	S	S
8	0,81	S	L
9	0,86	S	S
10	0,32	L	L
11	0,87	S	S
12	0,99	S	S
13	0,96	S	S
14	0,98	S	S
15	0,55	L	L
16	0,65	S	L
17	0,72	S	S
18	0,45	L	S
Threshold : 0,6			



Gambar 2. Perbandingan evaluasi performa SA, TIE, dan SA + TIE

5. Kesimpulan

Dalam penelitian ini telah diusulkan metode baru untuk membedakan antara *bot spammer* dan *legitimate user* dengan mengintegrasikan SA yang berdasarkan emosi dan TIE. S*entiment analysis* (SA) digunakan untuk mendeteksi ungkapan ekspresi ataupun opini yang terkandung dalam *tweet.* Sedangkan TIE digunakan untuk menangkap keteraturan waktu memposting *tweet* yang menunjukkan *tweet* diunggah secara otomatis.

Serangkaian percobaan telah dilakukan untuk mengevaluasi kinerja dari metode yang diusulkan. Dari hasil eksperimen dapat disimpulkan bahwa penggabungan SA dan TIE dapat mengoptimalkan performa sistem secara keseluruhan dalam mengidentifikasi *bot spammer* dengan nilai *precision* dan *recall* masing-masing 83% dan 91%.

Penelitian lebih lanjut dapat dilakukan untuk menyelidiki hubungan semantik pada *tweet* dimana kata dapat mengalami bias atau bermakna ganda. Selain itu dalam pendeteksian *bot spammer* dapat juga ditambahkan beberapa karakterisktik seperti jumlah *follower*, *following*, dan *retweet*. Beberapa *bot spammer* secara otomatis mengikuti *legitimate user* dan juga melakukan *retweet* secara otomatis terhadap *tweet* yang mencakup kata atau frase tertentu.

Referensi

Arifin, A. Z., & Setiono, A. N. 2002. Klasifikasi Dokumen Berita Kejadian Berbahasa Indonesia dengan Algoritma Single Pass Clustering. In *Prosiding Seminar on Intelligent*

Technology and its Applications (SITIA), Teknik Elektro, Institut Teknologi Sepuluh Nopember Surabaya.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. 2012. Detecting automation of twitter accounts: Are you a human, bot, or cyborg?. *Dependable and Secure Computing, IEEE Transactions on*, *9*(6), 811-824.

Drucker, H., Wu, D., & Vapnik, V. N. 1999. Support vector machines for spam categorization. *Neural Networks, IEEE Transactions on*, *10*(5), 1048-1054.

Esuli, A., & Sebastiani, F. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC* (Vol. 6, pp. 417-422).

Heron, S. 2009. Technologies for spam detection. *Network Security*, *2009*(1), 11-15.

Hu, M., & Liu, B. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.

Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. 2013. Ontology-based sentiment analysis of twitter posts. *Expert systems with applications*, *40*(10), 4065-4074.

Li, W., & Xu, H. 2014. Text-based emotion classification using emotion cause extraction. *Expert Systems with Applications*, *41*(4), 1742-1749.

Lima, A. C. E., de Castro, L. N., & Corchado, J. M. 2015. A polarity analysis framework for Twitter messages. *Applied Mathematics and Computation*, *270*, 756-767.

Liu, B. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167.

Miller, Z., Dickinson, B., Deitrick, W., Hu, W., & Wang, A. H. 2014. Twitter spammer detection using data stream clustering. *Information Sciences*, *260*, 64-73.

Mohammad, S. M., Zhu, X., Kiritchenko, S., & Martin, J. 2014. Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing & Management*. Elsevier.

Montejo-Ráez, A., Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. 2014. Ranked wordnet graph for sentiment polarity classification in twitter. *Computer Speech & Language*, *28*(1), 93-107.

Moraes, R., Valiati, J. F., & Neto, W. P. G. 2013. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, *40*(2), 621-633.

Perdana, R. S., Muliawati, T. H., & Alexandro, R. 2015. Bot Spammer Detection in Twitter Using Tweet Similarity And Time Interval Entropy. *Jurnal Ilmu Komputer dan Informasi*, *8*(1), 20-26.

Poria, S., Cambria, E., Winterstein, G., & Huang, G. B. 2014. Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowledge-Based Systems*, *69*, 45-63.

Takhteyev, Y., Gruzd, A., & Wellman, B. 2012. Geography of Twitter networks. *Social networks*, *34*(1), 73-81.

Tala, F. Z. 2003. A study of stemming effects on information retrieval in Bahasa Indonesia. *Institute for Logic, Language and Computation Universeit Van Amsterdam*.

Tausczik, Y. R., & Pennebaker, J. W.(2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, *29*(1), 24-54.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. 2010. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, *61*(12), 2544-2558.

Verma, M., & Sofat, S. 2014. Techniques to Detect Spammers in Twitter-A Survey. *International Journal of Computer Applications*, *85*(10), 27-32.

Yang, C., Harkreader, R. C., & Gu, G. 2011. Die free or live hard? empirical evaluation and new design for fighting evolving twitter spammers. In*Recent Advances in Intrusion Detection* (pp. 318-337). Springer Berlin Heidelberg.