Register : Jurnal Ilmiah Teknologi Sistem Informasi 5 (1) 2019 37-46
ISSN 2502-3357 (*Online*)
ISSN 2503-0477 (*Print*)

Tersedia *online* di www.journal.unipdu.ac.id
Unipdu
**Terakreditasi S2** – SK No. 34/E/KPT/2018
Halaman jurnal di www.journal.unipdu.ac.id/index.php/register

# Deteksi *Bot Spammer* Twitter Berbasis *Time Interval Entropy* dan *Global Vectors for Word Representations Tweet's Hashtag*

*Arif Mudi Priyatno [a], Muhammad Mirza Muttaqi [b], Fahmi Syuhada [c], Agus Zainal Arifin [d]*

[a,b,c,d] *Teknik Informatika, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*

*email: [a]arif.18051@mhs.its.ac.id, [b]mirza.18051@mhs.its.ac.id, [c]fahmi.18051@mhs.its.ac.id, [d]agus.za@its-sby.edu*

**INFO ARTIKEL**

***Sejarah artikel:***
Menerima 18 Desember 2018
Revisi 11 April 2019
Diterima 11 April 2019
*Online* 29 Mei 2019

***Kata kunci:***
*bot spammer*
CNN
Glove
*hashtag*
Twitter

***Keywords:***
*bot spammer*
CNN
Glove
*hashtag*
Twitter

***Style APA dalam menyitasi artikel ini:***
Priyatno, A. M., Muttaqi, M. M., Syuhada, F., & Arifin, A. Z. (2019). Deteksi Bot Spammer Twitter Berbasis Time Interval Entropy dan Global Vectors for Word Representations Tweet's Hashtag. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 5(1), 37-46.

**ABSTRAK**

*Bot spammer* merupakan penyalahgunaan *user* dalam menggunakan Twitter untuk menyebarkan pesan spam sesuai dengan keinginan *user*. Tujuan spam mencapai *trending* topik yang ingin dibuatnya. Penelitian ini mengusulkan deteksi *bot spammer* pada Twitter berbasis *Time Interval Entropy* dan *global vectors for word representations* (Glove). *Time Interval Entropy* digunakan untuk mengklasifikasi akun *bot* berdasarkan deret waktu pembuatan *tweet*. Glove digunakan untuk melihat *co-occurrence* kata *tweet* yang disertai *Hashtag* untuk proses klasifikasi menggunakan *Convolutional Neural Network* (CNN). Penelitian ini menggunakan data API Twitter dari 18 akun *bot* dan 14 akun legitimasi dengan 1.000 *tweet* per akunnya. Hasil terbaik *recall*, *precision*, dan *f-measure* yang didapatkan yaitu 100%; 100%, dan 100%. Hal ini membuktikan bahwa Glove dan *Time Interval Entropy* sukses mendeteksi *bot spammer* dengan sangat baik. *Hashtag* memiliki pengaruh untuk meningkatkan deteksi *bot spammer*.

***ABSTRACT***

*Spam spammers are users' misuse of using Twitter to spread spam messages in accordance with user wishes. The purpose of spam is to reach the required trending topic. This study proposes detection of bot spammers on Twitter based on Time Interval Entropy and global vectors for word representations (Glove). Time Interval Entropy is used to classify bot accounts based on the tweet's time series, while glove views the co-occurrence of tweet words with Hashtags for classification processes using the Convolutional Neural Network (CNN). This study uses Twitter API data from 18 bot accounts and 14 legitimacy accounts with 1000 tweets per account. The best results of recall, precision, and f-measure were 100%respectively. This proves that Glove and Time Interval Entropy successfully detects spams, with Hash tags able to increase the detection of bot spammers.*



## 1. Pendahuluan

Teknologi yang muncul melalui Jaringan Sosial *online* telah mengarah pada pengembangan berbagai *platform* dengan jutaan entitas sosial yang berkolaborasi dan berkomunikasi satu sama lain (Sedhai & Sun, 2018). Saat ini, Jaringan Sosial *Online* atau *Online Social Network* (OSN) telah menjadi bagian dari rutinitas sehari-hari masyarakat umum. Para pengguna dapat menghabiskan banyak waktu di OSN populer tempat mereka menyimpan dan menyebarkan informasi pribadi. Di antara berbagai jenis OSN, Twitter dianggap sebagai salah satu OSN paling populer (Perdana, Muliawati, & Alexandro, 2015). Twitter merupakan situs web media sosial mikro *blogging*. Perbedaan Twitter yang menonjol dengan OSN lain, seperti Facebook dan LinkedIn, yaitu Twitter membatasi *posting* atau *tweets* (pesan berbasis teks) hanya 280 karakter. Twitter juga unik dalam hubungan yang dapat diarahkan, sedangkan pada Facebook sebagian besar hubungan bersifat dua arah (Fields, 2016).

A. M. Priyatno dkk./Register 5 (1) 2019 37-46
ISSN 2502-3357 (*Online*)
ISSN 2503-0477 (*Print*)

Pengguna Twitter diidentifikasi dengan nama pengguna mereka secara opsional dengan nama asli mereka. Pengguna "A" mulai mengikuti pengguna lain dan *tweet* mereka akan muncul di halaman A. Pengguna A dapat diikuti kembali jika pengguna lain menginginkannya. Topik yang sedang tren di Twitter dapat diidentifikasi dengan *hashtag* atau tanda pagar (tagar) "#" (Daffa, Bamasag, & AlMansour, 2018). Twitter memiliki manfaat untuk perusahaan dan organisasi. Twitter, sebagai saluran yang efektif untuk terhubung dengan pelanggan, mempromosikan atau menjual produk. Dengan keunggulan ini, Twitter semakin banyak digunakan untuk penyebaran informasi berskala besar di berbagai bidang kehidupan manusia, seperti pemasaran, jurnalisme, atau hubungan masyarakat (Nguyen & Takeda, 2016).

Popularitas Twitter telah menarik banyak pengguna Twitter untuk menyebarkan banyak pesan spam (Perdana, Muliawati, & Alexandro, 2015). Spam merupakan suatu pesan yang dikirimkan oleh seorang pengguna yang diistilahkan *spammer* ke pengguna lain secara bertubi-tubi tanpa dikehendaki oleh targetnya. Selain spam, pengiriman pesan juga dapat berbentuk suatu tulisan lini masa maupun komentar yang berfungsi memviralkan sebuah topik tertentu. *Spammer* biasanya merupakan akun komputer yang dibuat seseorang dengan suatu tujuan dan maksud tertentu. Untuk mendapatkan jangkauan yang lebih luas kepada calon korban, *spammer* dikenal berteman (atau mengikuti terminologi Twitter) pengguna yang tidak terkait, mengirim pesan yang tidak diinginkan dan menyamarkan komponen berbahaya (misalnya, menggunakan *shortener* URL untuk mengganti URL *malicious appearing*) (Amleshwaram, Reddy, Yadav, Gu, & Yang, 2013).

Jenis spam yang paling dikenal di Twitter adalah untuk menangkap topik yang sedang tren (Martinez-Romo & Araujo, 2013). Setiap kali peristiwa yang patut dicatat terjadi, pengguna mencoba untuk mengekspresikan pendapat mereka atau berbagi informasi tentang acara tersebut menggunakan *hashtag*. Jika topiknya paling banyak diciak (*tweet*) di hari itu, terlihat oleh semua pengguna Twitter di beranda Twitter sebagai topik yang sedang tren. *Spammer* menggunakan tagar yang sama agar dapat dilihat oleh basis pengguna yang besar, agar *user* Twitter lain mengikuti peristiwa yang sedang tren tetapi dengan URL yang tidak diminta mengarah ke situs web yang tidak terkait (Bindu, Mishra, & Thilagam, 2018).

Program otomatis atau lebih dikenal sebagai *bot*, kependekan dari *robot*, tidak memerlukan operator manusia untuk melaksanakan tugasnya. Studi awal menunjukkan bahwa sebagian besar pesan spam di Twitter dihasilkan secara otomatis oleh *bot* (Amleshwaram, Reddy, Yadav, Gu, & Yang, 2013) dan hanya sedikit dari mereka yang dipos (*post*) secara manual oleh manusia (Chu, Gianvecchio, Wang, & Jajodia, 2012). *Spammer bot* dapat secara otomatis menghasilkan pesan spam pada waktu interval tertentu menggunakan penjadwal pekerjaan (Zhang & Paxson, 2011). Penggunaan *bot* dapat mengurangi biaya tinggi mengelola akun spam secara manual, sehingga lebih mudah bagi *spammer* untuk menghasilkan lebih banyak pesan spam di Twitter.

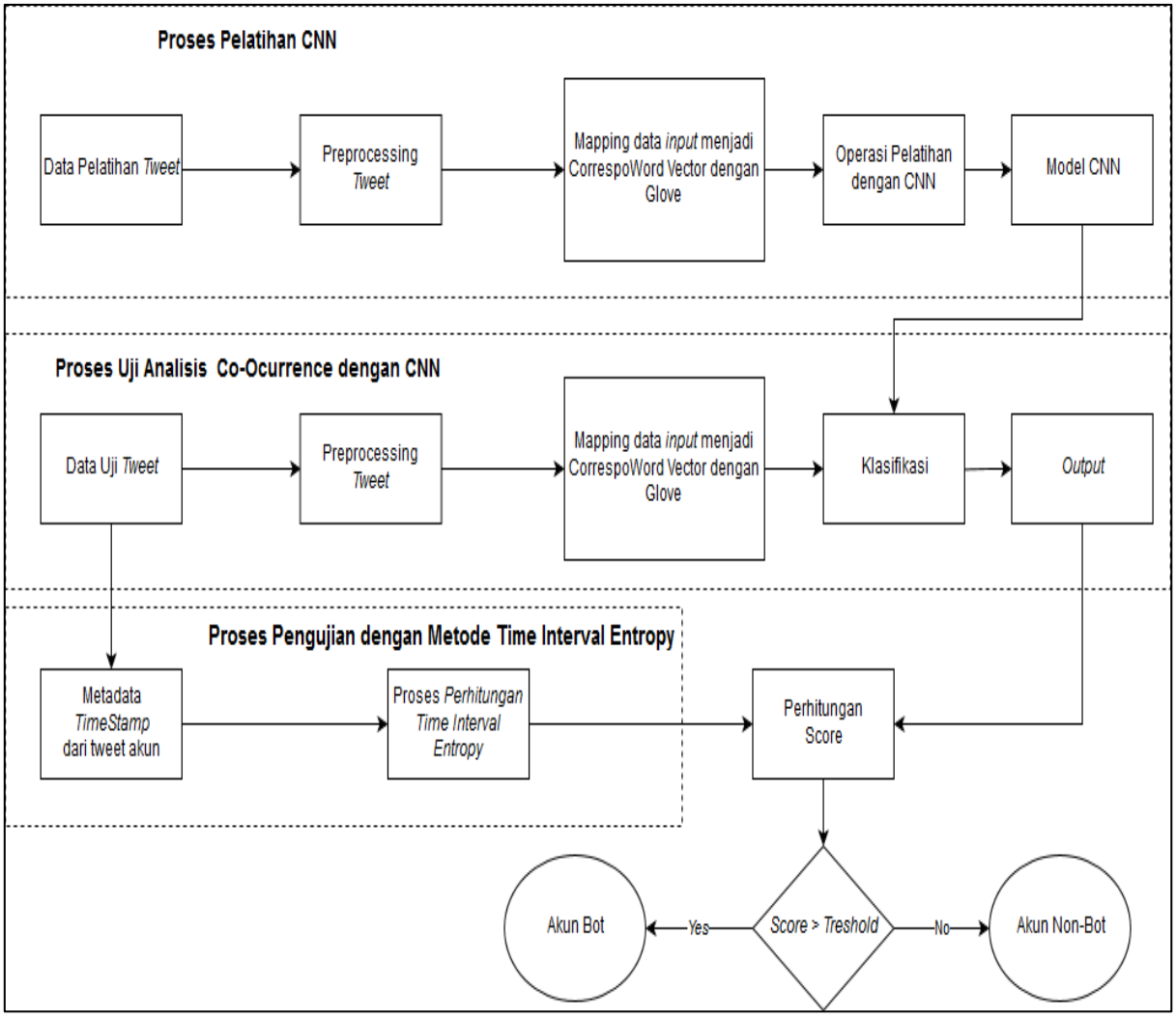
Jumlah pesan spam yang meningkat dapat memperburuk pengalaman pengguna Twitter. Hal ini mencemari informasi di Twitter dan membuang sumber daya dari pengguna (Kuzi, Shtok, & Kurland, 2016). Oleh karena itu, dibutuhkan upaya yang lebih untuk menghentikan pengembangan dari *spammer* di Twitter. Twitter sebenarnya telah menyediakan mekanisme untuk menghentikan pengembangan spam dengan mengundang para pengguna untuk aktif melaporkan pesan dan akun spam. Akan tetapi, ini membutuhkan banyak waktu dan sumber daya. Sementara, apabila salah memberikan label akun pengguna yang sah sebagai spam, dapat membahayakan ketergantungan pengguna terhadap Twitter.

Suatu akun dapat dikatakan sebagai akun *bot* berdasarkan beberapa faktor. Pertama, faktor waktu pembuatan setiap *tweet*. Umumnya, pembuatan *tweet* oleh akun *bot* dilakukan dengan waktu yang terjadwal (Daffa, Bamasag, & AlMansour, 2018). Hal ini disebabkan akun *bot* merupakan *script* program yang berjalan otomatis sesuai dengan perintah yang diterimanya. Kedua, akun *bot* memiliki ciri-ciri yaitu kecenderungan *tweet* yang satu dengan yang lainnya tidak berbeda jauh. *Tweet* tersebut memiliki target yang ingin dicapai seperti *trending* topik atau menyampaikan suatu pesan agar diketahui oleh banyak pihak. Proses mencapai tujuan biasanya selalu diselingi *hashtag* sebagai tanda atau kode yang mengarahkan kepada *tweet* yang dibuatnya. Pada umumnya, *bot spammer* melakukan *tweet* dengan membubuhkan *hashtag*. *Hashtag* memiliki tujuan tertentu yang diinginkan. Analisis *co-occurence* atau kedekatan terhadap setiap *tweet* dengan faktor kedua ini dapat dilakukan untuk

ISSN 2502-3357 (*Online*)
A. M. Priyatno dkk./Register 5 (1) 2019 37-46 ISSN 2503-0477 (*Print*)

mendeteksi *bot* akun (Aditya, Hani'ah, Fitrawan, Arifin, & Purwitasari, 2016). Setiap *tweet* akan melewati proses klasifikasi dengan *machine learning* untuk ditentukan apakah akun yang membuat *tweet* tersebut termasuk *bot* akun atau tidak. Salah sata *machine learning* yang saat ini sedang banyak dikembangkan yaitu *Convolutional Neural Network* (Schmidhuber, 2015).

Penelitian ini mengusulkan deteksi *bot spammer* pada Twitter berbasis *Time Interval Entropy* dan *global vectors for word representations* (Glove). *Time Interval Entropy* digunakan untuk menganalisis tingkat keteraturan dari deret waktu pembuatan *tweet*. Sementara, Glove digunakan pada proses analisis *co-ocurrence tweet* untuk dapat masuk ke dalam CNN. Strategi ini diharapkan mampu untuk mendeteksi *bot spammer*, sehingga mengurangi *user bot spammer* yang berkeliaran pada jejaring sosial media Twitter.




Gambar 1. Diagram blok kombinasi metode yang diusulkan

## 2. Data dan Metode Penelitian

### *2.1. Data*

Dalam penelitian ini, peneliti mengumpulkan data Twitter untuk mendeteksi *bot* atau bukan dengan memanfaatkan API yang sudah disediakan oleh Twitter. Terdapat 32 akun, setiap akun diambil maksimal 1.000 *tweet* dengan rentang waktu 2 tahun terakhir. Data yang didapatkan berupa metadata dari setiap *tweet* berdasarkan akun yang diambil. Metadata yang digunakan pada penelitian ini adalah isi dan *timestamp* atau data waktu dari *tweet*. Pada proses pengambilan data akun Twitter, 18 akun *bot* ditentukan *secbot* dan 14 akun legitimasi atau akun *nonbot*. Data yang diperoleh dari Twitter akan dimasukkan ke dalam sebuah *database* untuk dilakukan proses pendeteksian *bot* atau bukan.

Karakteristik sebuah akun dapat dikatakan sebagai akun *bot*, yaitu terdapat beberapa ciri yang dapat dideteksi secara manual (Yang, Harkreader, & Gu, 2011): (1) kiriman berupa *link* menuju ke *website* tertentu; (2) kiriman dengan ditambahkan foto dalam *website* agar lebih menarik pembaca untuk

ISSN 2502-3357 (*Online*)
A. M. Priyatno dkk./Register 5 (1) 2019 37-46 ISSN 2503-0477 (*Print*)

mengakses *link* tersebut; (3) kiriman satu dengan yang lain tidak jauh berbeda; (4) kiriman yang sedikit mendapatkan *like*, komentar, atau *retweet* karena kurang menarik; dan (5) kiriman yang banyak menggunakan *hashtag* paling populer untuk mendapatkan pembaca.

### 2.2. *Metode*

Pada penelitian ini diusulkan sebuah kombinasi metode baru untuk membedakan antara akun *bot* dengan akun legitimasi. Metode pertama yang diintegrasikan yaitu metode *Convolutional Neural Network* (CNN) yang digunakan pada proses klasifikasi akun berdasarkan pemrosesan *tweet*. Metode yang kedua yaitu penentuan akun *bot* berdasarkan *Time Interval Entropy tweet* dari sebuah akun yang yang diteliti. Gambar 1 merepresentasikan skema usulan kontribusi penelitian yang peneliti lakukan.

Pengaplikasian metode CNN dibagi menjadi dua proses, yaitu proses *training* dan proses *testing*. Pada proses *training* dibentuk suatu model CNN dari data *tweet* pelatihan yang didapatkan. Sebelum masuk ke dalam pembentukan model, terlebih dahulu data dilakukan *preprocessing* dan *mapping*.

*Preprocessing* merupakan proses pembersihanan isi dari setiap *tweet* yang masuk ke proses selanjutnya. Adapun proses yang dilakukan yaitu mengganti huruf kapital menjadi nonkapital, menghapus karakter non-ASCI selain karakter *hashtag* atau tanda pagar "#", dan *stemming tweet* untuk mengembalikan kata-kata pada *tweet* ke bentuk kata dasarnya. Pada penelitian ini, kami menggunakan *framework* Sastrawi yang tersedia dalam bahasa Python untuk melakukan *preprocessing* (Aditya, Hani'ah, Fitrawan, Arifin, & Purwitasari, 2016). Setelah dilakukan *preprocessing*, selanjutnya dilakukan *mapping*.

*Mapping* digunakan untuk konversi *word* ke *vector* yang disebut *data word embedding*. *Data word embedding* yang dihasilkan akan masuk ke proses pembentukan model CNN. Proses pengujian data *test tweet* data dilakukan setelah pembentukan model selesai dilakukan. Klasifikasi data dilakukan menggunakan model yang telah dibentuk. Sebelum masuk ke proses klasifikasi, dilakukan konversi menjadi *word embedding* seperti pada proses pelatihan. Hasil dari klasifikasi akan menghasilkan sebuah *output* yang dikombinasikan dengan hasil dari metode *Time Interval Entropy*.

### 2.3. *Time Interval Entropy*

Komponen entropi mendeteksi waktu berkala atau reguler dari pesan yang dikirim oleh seorang pengguna media sosial. Nilai entropi waktu pada suatu rentetan *tweet* menunjukkan bahwa terjadi suatu perilaku pembuatan *tweet* yang teratur. Hal ini memunculkan suatu proses otomatisasi pembuatan suatu *tweet* yang kemungkinan dilakukan oleh akun *bot*. Nilai *entropy* tinggi pada suatu rentetan *tweet* yang dibuat seorang pengguna menunjukkan ketidakteraturan suatu pembuatan *tweet*. Hal ini dapat dimungkinkan bahwa *tweet* tersebut dibuat oleh pengguna manusia asli (Chu, Gianvecchio, Wang, & Jajodia, 2012).

Setiap kiriman atau *tweet* pada akun media sosial seperti Twitter memiliki metadata *timestamp*. Antara sebuah *tweet* dengan *tweet* yang lainnya memiliki interval waktu pengunggahan. *Time Interval Entropy* merupakan teknik yang digunakan untuk menganalisis pola keteraturan dari interval waktu. Persamaan 1 dan Persamaan 2 digunakan untuk menganalisis pola dari interval waktu dari kumpulan *tweet*.

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

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

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

$\boldsymbol{\Delta T}$ merepresentasikan interval waktu antara *tweet*, dengan $\boldsymbol{P\Delta T(t_i)}$ menunjukkan probabilitas dari interval waktu $\boldsymbol{\Delta T_i}$. Pada Persamaan 1 direpresentasikan proses deteksi *bot* dengan analisis rentang waktu ($\boldsymbol{T}$) dari deret waktu pembuatan setiap *tweet* pada akun. Nilai *entropy* ($\boldsymbol{P\Delta T}$) didapatkan dari deret waktu tersebut menggunakan Persamaan 1 dan Persamaan 2. Jika *entropy* yang didapatkan lebih kecil dari nilai ambang atau *threshold* = 0,5, disimpulkan bahwa akun tersebut merupakan akun *bot*. Nilai *threshold* didapatkan dari hasil percobaan menggunakan akun data pelatihan. Persamaan 3 digunakan untuk menggabungkan antara *co-occurance* dan *Time Interval Entropy*. Variabel $\boldsymbol{k}$ merupakan

ISSN 2502-3357 (*Online*)
A. M. Priyatno dkk./Register 5 (1) 2019 37-46 ISSN 2503-0477 (*Print*)

akun dari Twitter. $\boldsymbol{Score_k}$ merupakan nilai interval waktu yang didapatkan setiap akun. Variabel $\boldsymbol{\alpha}$ dan $\boldsymbol{\beta}$ menunjukan nilai bobot dari nilai *co-occurance* dan *Time Interval Entropy*. Jumlah dari $\boldsymbol{\alpha}$ dan $\boldsymbol{\beta}$ harus sama dengan 1. $\boldsymbol{H_k}$ adalah *Time Interval Entropy* akun. *Cooccurance* adalah nilai *co-occurance* akun.

### 2.4. *Global Vectors for Word Representations (Glove)*

Glove adalah algoritma untuk mendapatkan representasi vektor dari kata-kata dengan cara melibatkan seluruh informasi yang telah diperoleh dari setiap *tweet*. Representasi vektor *tweet* tersebut dinamakan *word embedding*. *Global vectors for word representations* (Glove) merupakan proses pembentukan *word co-occurrence matrix* dari suatu kata. Proses ini merupakan proses terbaik untuk menghasilkan *word embeddings*. Hal ini terbukti dari tingkat keberhasilannya yang mencapai 75% untuk tes analogi kata (Pennington, Socher, & Christopher, 2014). Teknik mendapatkan *word embeddings* dibagi dua metode. Metode pertama yaitu faktorisasi matrik, dengan *word embeddings* matrik dibuat berdasarkan jumlah kemunculan kata, kemudian dikonversikan ke dalam vektor berdimensi tertentu. Proses selanjutnya, *context window,* yaitu proses untuk membandingkan antarkata yang sering muncul pada setiap *tweet* yang akan dibandingkan.

### 2.5. *Convolutional Neural Networks (CNN)*

*Convolutional Neural Networks* (CNN) pertama kali ditemukan pada kasus penyelesaian data gambar. Namun, CNN bisa juga diterapkan pada kasus *Natural Language Processing* (NLP), serta CNN memberikan hasil yang bagus. Proses CNN pada data *text* (teks) secara umum sama dengan proses CNN pada data gambar. Ada perbedaan antara proses tersebut, yaitu terletak pada jumlah *layer* konvolusi dan layer lainnya. Pada kasus data gambar jumlah *layer* bisa sesuai dengan dimensinya, bisa dua dimensi atau bahkan tiga dimensi. Sementara, pada kasus teks, jumlah lapisannya sebanyak satu *layer* atau dimensi. *Layer* konvolusi yang satu dimensi menerima masukan satu dimensi yang merupakan representasi dari dokumen yang dimasukkan. Pada tahap ini, yang didapatkan yaitu fitur yang berpengaruh pada proses klasifikasi.

Pada klasifikasi dokumen teks dengan Glove, pembentukan vektor dokumen dilakukan secara merata untuk seluruh dokumen. Penelitian Kenter, Borisov, and Rijke (2016) menyatakan hasil yang efektif dalam melakukan representasi vektor dokumen. Dua model CNN pada penelitian ini yaitu model konvolusi dan klasifikasi. Fungsi aktivasi yang digunakan yaitu fungsi aktivasi *Rectified Linear Unit* (ReLU) pada *layer* konvolusi dan *layer* tersembunyi pada klasifikasi.

### 2.6. *Evaluasi Kinerja*

Evaluasi kinerja dari metode yang diusulkan menggunakan konvolusi matrik dengan teknik *recall*, *precision,* dan *f-measure*. Teknik ini digunakan untuk menghitung secara kuantitatif. *Recall* adalah jumlah dokumen yang relevan yang ditemukan pada hasil prediksi. *Precision* adalah jumlah *tweet* relevan yang didapatkan dari total jumlah hasil deteksi yang dilakukan. *F-measure* adalah akurasi dengan mempertimbangkan *precision* dan *recall*. Persamaan 4 adalah *recall*, Persamaan 5 adalah *precision,* dan Persamaan 5 adalah *f-measure*.

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$Precision = \frac{TP}{TP+TN} \tag{5}$$

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision+Recall} \tag{6}$$

Perhitungan *recall* dan *precision* menggunakan *True Positive* (TP), *False Positive* (FP), dan *False Negative* (FN). TP merupakan data akun *bot spammer* yang benar dideteksi. FP merupakan data akun *bot spammer* yang salah dideteksi. FN merupakan data akun legitimasi yang gagal dideteksi.

## 3. Hasil dan Pembahasan

Nilai *threshold* pada penelitian ini ditentukan berbeda-beda setiap skenario pengujian. Hal ini berdasarkan pada percobaan-percobaan yang dilakukan untuk mendapatkan *recall* dan *precision* yang terbaik. Skenario uji coba pertama melihat *co-occurrence* antara kata pada *tweet* dengan menghilangkan *hashtag*, *mention,* dan juga *link* web jika ada. Skenario uji coba kedua melihat *co-occurrence* antarkata pada *tweet* dengan tidak menghilangkan *hashtag* yang digunakannya. Skenario uji coba ketiga melihat waktu

ISSN 2502-3357 (*Online*)
ISSN 2503-0477 (*Print*)

interval *tweet* yang dilakukan dengan *Time Interval Entropy*. Skenario uji coba keempat menggabungkan skenario uji coba pertama dan skenario uji coba ketiga. Skenario uji coba kelima menggabungkan skenario uji coba kedua dengan skenario uji coba ketiga.

Tabel 1. Hasil klasifikasi Glove tanpa *hashtag*

| No | *Score* Akun | Jumlah *tweet* | L | Jumlah *Tweet Bot* | Prediksi |
|---|---|---|---|---|---|
| 1 | 0,135 | 1000 | 135 | 865 | B |
| 2 | 0,123 | 1000 | 123 | 877 | B |
| 3 | 0,413 | 995 | 411 | 584 | B |
| 4 | 0,353 | 999 | 353 | 646 | B |
| 5 | 0,159 | 1000 | 159 | 841 | B |
| 6 | 0,152 | 1000 | 152 | 848 | B |
| 7 | 0,185 | 1000 | 185 | 815 | B |
| 8 | 0,116 | 1000 | 116 | 884 | B |
| 9 | 0,144 | 1000 | 144 | 856 | B |
| 10 | 0,67 | 1000 | 670 | 330 | L |
| 11 | 0,823 | 999 | 822 | 177 | L |
| 12 | 0,66 | 1000 | 660 | 340 | L |
| 13 | 0,791 | 999 | 790 | 209 | L |
| 14 | 0,855 | 998 | 853 | 145 | L |
| 15 | 0,576 | 100 | 576 | 424 | L |
| 16 | 0,793 | 1000 | 793 | 207 | L |

*Threshold* = 0,5; B = *Bot Spammer*; L = Legitimasi

Tabel 2. Hasil klasifikasi Glove dengan *hashtag*

| No | *Score* Akun | Jumlah *tweet* | L | Jumlah *Tweet Bot* | Prediksi |
|---|---|---|---|---|---|
| 1 | 0,092 | 1000 | 92 | 908 | B |
| 2 | 0,065 | 1000 | 65 | 935 | B |
| 3 | 0,347 | 995 | 345 | 650 | B |
| 4 | 0,335 | 999 | 335 | 664 | B |
| 5 | 0,093 | 1000 | 93 | 907 | B |
| 6 | 0,069 | 1000 | 69 | 931 | B |
| 7 | 0,136 | 1000 | 136 | 864 | B |
| 8 | 0,058 | 1000 | 58 | 942 | B |
| 9 | 0,109 | 1000 | 109 | 891 | B |
| 10 | 0,68 | 1000 | 680 | 320 | L |
| 11 | 0,785 | 999 | 784 | 215 | L |
| 12 | 0,628 | 1000 | 628 | 372 | L |
| 13 | 0,757 | 999 | 756 | 243 | L |
| 14 | 0,829 | 998 | 827 | 171 | L |
| 15 | 0,612 | 1000 | 612 | 388 | L |
| 16 | 0,68 | 1000 | 680 | 320 | L |

*Threshold* = 0,5; B = *Bot Spammer*; L = Legitimasi

Pada skenario uji coba pertama menggunakan Glove tanpa menggunakan *hashtag*. Jika *tweet user* lebih kecil dari *threshold* 0,5, akan dikategorikan sebagai *bot*. Hasil klasifikasi *bot* dan legitimasi semua *tweet* pada Twitter diambil rata-rata setiap akunnya. Jika akun tersebut lebih kecil dari *threshold*, akun tersebut dikategorikan ke dalam akun *bot spammer*. Hasil dari skenario uji coba pertama dapat dilihat pada Tabel 1. Pada setiap akun dapat kita lihat masih ada *tweet* yang dikategorikan *bot* dan legitimasi. Hal ini disebabkan *co-occurrence* kata pada *tweet* memiliki kedekatannya masing-masing. Kata *tweet* ada yang dekat dengan ciri-ciri *bot spammer* dan juga ada yang dekat dengan legitimasi. Akun *bot spammer* secara keseluruhan memiliki nilai yang kecil atau di bawah *threshold*, sehingga dikategorikan ke dalam *bot spammer* sesuai dengan labelnya. Hal ini disebabkan ciri-ciri akun *bot spammer* untuk rata-rata keseluruhan *tweet* yang dilakukannya memiliki *co-occurrence* yang sama. Sebagai contoh, pada akun pertama jumlah *tweet* yang dimiliki 1.000 *tweet* dan *tweet* terdeteksi *bot* sebanyak 865. *Recall*, *precision*, dan *f-measure tweet* secara berurutan mendapatkan sebesar 79,75%; 80,23%; dan 79,99%. Sementara,


http://doi.org/10.26594/register.v5i1.1382

ISSN 2502-3357 (*Online*)
A. M. Priyatno dkk./Register 5 (1) 2019 37-46 ISSN 2503-0477 (*Print*)

*recall*, *precision*, dan *f-measure* akun secara berurutan mendapatkan nilai sebesar 100%; 100%; dan 100%. Skenario ini memiliki kelemahan, yaitu arah *score* yang dihasilkan mengarahkan kepada hasil di bawah *threshold*, sehingga lebih mudah untuk dikenali sebagai *bot spammer*. Hal ini memberikan efek bahwa apabila suatu akun legitimasi memiliki kedekatan *tweet* antara satu dengan yang lainnya akan terdeteksi sebagai *bot spammer*.

Skenario uji coba kedua menggunakan Glove dengan memasukkan *hashtag* sebagai fiturnya. Hal ini dilakukan karena pada setiap *bot* cenderung memiliki kesamaan satu dengan yang lainnya. Skenario ini menggunakan *threshold* 0,5 untuk *tweet* dan akun. Jika hasil *tweet* nilai akun lebih kecil dari *threshold*, dikategorikan *bot spammer*. Hasil skenario kedua dapat dilihat pada Tabel 2. Hasil dari skenario ini yaitu meningkatkan jumlah *tweet* yang dianggap sebagai *tweet bot*. Hal ini disebabkan *hashtag* yang terdapat pada setiap *tweet* memiliki kedekatan yang sangat dekat, bahkan bisa saja sama. Kelemahan yang terlihat jika menggunakan *hashtag* sebagai salah satu hal penting adalah akun legitimasi terlalu banyak menggunakan *hashtag* dalam setiap *tweet* yang dilakukan. Maka, *score* yang dihasilkan akan memperbesar kemungkinan akun tersebut menuju *bot spammer*. Selain itu, *tweet* yang memiliki kemiripan antara satu *tweet* dengan *tweet* lainnya mampu membuat hasilnya menjadi *bot spammer*. Nilai *recall*, *precision*, dan *f-measure tweet* secara berurutan mendapatkan 79,13%; 85,52%; dan 82,20%. Nilai *recall*, *precision*, dan *f-measure* akun secara berurutan mendapatkan 100%; 100%; dan 100%.

Tabel 3. Hasil klasifikasi *Time Interval Entropy*

| No | *Score* Akun | Prediksi | Label |
|---|---|---|---|
| 1 | 3,0273 | B | B |
| 2 | 3,4204 | B | B |
| 3 | 2,9203 | B | B |
| 4 | 6,4439 | L | B |
| 5 | 2,9360 | B | B |
| 6 | 3,2627 | B | B |
| 7 | 5,4599 | L | B |
| 8 | 3,4982 | B | B |
| 9 | 3,9787 | B | B |
| 10 | 8,0764 | L | L |
| 11 | 6,6914 | L | L |
| 12 | 6,9940 | L | L |
| 13 | 5,7276 | L | L |
| 14 | 4,9381 | L | L |
| 15 | 7,5625 | L | L |
| 16 | 6,1511 | L | L |

*Threshold* = 0,5; B = *Bot Spammer*; L = Legitimasi

Tabel 4. Hasil klasifikasi Glove tanpa *hashtag* dan *Time Interval Entropy*

| No | *Score* Akun | Prediksi | Label |
|---|---|---|---|
| 1 | 0,26903 | B | B |
| 2 | 0,28711 | B | B |
| 3 | 0,42095 | B | B |
| 4 | 0,61032 | B | B |
| 5 | 0,27694 | B | B |
| 6 | 0,29367 | B | B |
| 7 | 0,45183 | B | B |
| 8 | 0,28805 | B | B |
| 9 | 0,33450 | B | B |
| 10 | 0,89458 | L | L |
| 11 | 0,89396 | L | L |
| 12 | 0,82022 | L | L |
| 13 | 0,81456 | L | L |
| 14 | 0,80098 | L | L |
| 15 | 0,80833 | L | L |
| 16 | 0,84268 | L | L |

*Threshold* = 0,5; B = *Bot Spammer*; L = Legitimasi

ISSN 2502-3357 (*Online*)
A. M. Priyatno dkk./Register 5 (1) 2019 37-46 ISSN 2503-0477 (*Print*)

Skenario uji coba ketiga menggunakan *Time Interval Entropy*. Skenario ini melihat kebiasaan *bot* yang biasanya dilakukan *tweet* secara terjadwal. Hasil dari *Time Interval Entropy* dapat dilihat pada Tabel 3. Permasalahan yang muncul apabila menggunakan *Time Interval Entropy* yaitu apabila akun tersebut memiliki interval waktu *tweet* yang sangat kecil atau seperti otomatis pada akun legitimasi, akan ditandai sebagai akun *bot spammer*. Permasalahan lainnya jika menggunakan *Time Interval Entropy* yaitu apabila *user* legitimasi memiliki waktu yang seperti contoh terjadwal pada hari tertentu dan jam yang sama, akan mengakibatkan *user* legitimasi tersebut dikategorikan sebagai *bot spammer*. Hasil dari *recall*, *precision,* dan *f-measure* secara berurutan mendapatkan 100%; 77,78%; dan 82,35%.

Skenario uji coba keempat menggabungkan Glove tanpa *hashtag* dan *Time Interval Entropy* dengan rasio perbandingan bobot $\alpha$ dan $\beta$ yaitu 1:1. Hasil skenario uji coba keempat dapat dilihat pada Tabel 4. *Threshold* yang digunakan pada skenario ini adalah 0,65. Nilai akun yang didapatkan pada skenario uji coba keempat mengalami peningkatan daripada skenario 1. Hal ini menunjukan *Time Interval Entropy* memiliki pengaruh cukup besar jika dilakukan penggabungan keduanya. *Time Interval Entropy* membuat akun mendekati legitimasi atau meningkatkan nilai akun tersebut. Hasil nilai dari *recall*, *precision* ,dan *f-measure* secara berurutan mendapatkan 100%, 100%, dan 100%.

Tabel 5. Hasil klasifikasi Glove dengan *hashtag* dan *Time Interval Entropy*

| No | Score Akun | Prediksi | Label |
|---|---|---|---|
| 1 | 0,24818 | B | B |
| 2 | 0,25784 | B | B |
| 3 | 0,38887 | B | B |
| 4 | 0,60910 | B | B |
| 5 | 0,24288 | B | B |
| 6 | 0,25001 | B | B |
| 7 | 0,43026 | B | B |
| 8 | 0,25880 | B | B |
| 9 | 0,31927 | B | B |
| 10 | 0,91387 | L | L |
| 11 | 0,88542 | L | L |
| 12 | 0,81407 | L | L |
| 13 | 0,80714 | L | L |
| 14 | 0,79798 | L | L |
| 15 | 0,84139 | L | L |
| 16 | 0,78993 | L | L |

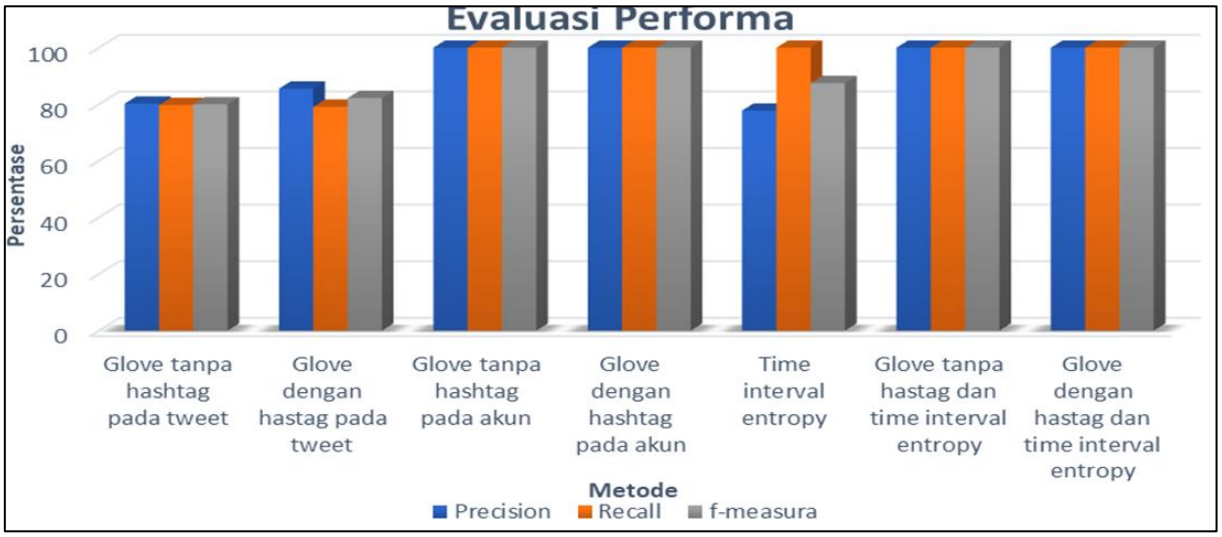
*Threshold* = 0,5; B = *Bot Spammer*; L = Legitimasi

Skenario uji coba kelima menggabungkan Glove dengan *hashtag* dan *Time Interval Entropy*. Rasio bobot $\alpha$ dan $\beta$ masih sama dengan skenario uji coba keempat, yaitu 1:1. *Threshold* yang digunakan yaitu 0,65. Hasil dari skenario uji coba kelima dapat dilihat pada Tabel 5. Nilai akun yang didapatkan dari skenario uji coba kelima mengalami kenaikan daripada skenario uji coba kedua. Hasil ini mengalami penurunan daripada skenario uji coba keempat. Sehingga dapat ditarik informasi bahwa penggunaan *hashtag* pada Glove memiliki pengaruh untuk membuat suatu akun legitimasi menjadi akun *bot spammer*. Akan tetapi, pada skenario ini dapat diperbaiki kelemahannya jika hanya menggunakan Glove dengan *hashtag* seperti yang terlihat pada skenario uji coba kedua, yaitu Tabel 2. Seperti yang terlihat pada *user* pertama di Tabel 2, *user* memiliki *score* 0,092, sedangkan pada skenario uji coba kelima ini yang terlihat pada Tabel 5, *user* pertama memiliki *score* 0,24818.

Gambar 2 menunjukan hasil evaluasi yang telah dilaksanakan. Glove tanpa *hashtag* pada *tweet* merupakan grafik persentase dari Tabel 1 pada setiap *tweet user* dengan nilai *recall*, *precision,* dan *f-measure* berurutan 79,75%; 80,23%; dan 79,99%. Glove dengan *hashtag* pada *tweet* merupakan grafik persentase dari Tabel 2 pada setiap *tweet user* dengan nilai *recall*, *precision,* dan *f-measure* secara berurutan 79,13%; 85,52%; dan 82,20%. Glove tanpa *hashtag* pada akun merupakan grafik Tabel 1 pada setiap akun *user* dengan nilai *recall*, *precision,* dan *f-measure* secara berurutan 100%; 100%; dan 100%. Glove dengan *hashtag* pada akun merupakan grafik dari Tabel 2 pada setiap *user* memiliki nilai *recall*, *precision,* dan *f-measure* secara berurutan 100%; 100%; dan 100%. *Time Interval Entropy* merupakan grafik dari Tabel 3 dengan nilai *recall*, *precision,* dan *f-measure* secara berurutan 100%; 77,78%; dan 82,35%. Glove tanpa

ISSN 2502-3357 (*Online*)
A. M. Priyatno dkk./Register 5 (1) 2019 37-46 ISSN 2503-0477 (*Print*)

*hashtag* dan *Time Interval Entropy* merupakan grafik dari Tabel 4 dengan nilai *recall*, *precision,* dan *f-measure* secara berurutan 100%; 100%; dan 100%. Glove dengan *hashtag* dan *Time Interval Entropy* merupakan grafik dari Tabel 5 dengan nilai *recall*, *precision,* dan *f-measure* secara berurutan 100%; 100%; dan 100%.




Gambar 2. Evaluasi perfoma skenario uji coba

## 4. Kesimpulan

Penelitian ini mengusulkan deteksi *bot spammer* pada Twitter berbasis *Time Interval Entropy* dan Glove. *Time Interval Entropy* digunakan untuk mengetahui keteraturan *tweet* yang dilakukan oleh *user*. Glove digunakan untuk mengetahui tingkat *co-occurrence* kata pada setiap *tweet* yang dibuatnya. Kedua metode tersebut digunakan untuk mendeteksi akun *bot spammer* dan legitimasi. Pada beberapa percobaan telah dilaksanakan evaluasi kinerja dari metode yang diusulkan. Hasil dari percobaan yang dilakukan bahwa dengan menggabungkan Glove dan *Time Interval Entropy* mampu untuk mendeteksi *bot spammer* dengan sangat baik. *Co-occurrence* kata pada setiap *tweet* memiliki pengaruh dalam deteksi *bot spammer*. *Hashtag* yang merupakan ciri-ciri dari *bot spammer* dapat dibuktikan dengan memiliki pengaruhnya dalam penentuan *bot spammer*. Nilai *recall*, *precision,* dan *f-measure* metode *Time Interval Entropy* dan Glove kata *tweet* dengan pengaruh *hashtag* secara berurutan adalah 100%; 100%; dan 100%. Hal ini membuktikan bahwa Glove dan *Time Interval Entropy* sukses mendeteksi *bot spammer* dengan sangat baik. *Hashtag* memiliki pengaruh untuk meningkatkan deteksi *bot spammer*. Pengembangan selanjutnya yang menantang yaitu jika *bot spammer* tersebut menggunakan gambar atau foto setiap *tweet* agar tidak terlihat *co-occurance* setiap *tweet* yang dilakukannya.

## 7. Referensi

Aditya, h. S., Hani'ah, M., Fitrawan, A. A., Arifin, A. Z., & Purwitasari, D. (2016). Deteksi Bot Spammer pada Twitter Berbasis Sentiment Analysis dan Time Interval Entropy. *Jurnal Buana Informatika, 7*(3).

Amleshwaram, A. A., Reddy, N., Yadav, S., Gu, G., & Yang, C. (2013). CATS: Characterizing automation of Twitter spammers. *2013 Fifth International Conference on Communication Systems and Networks (COMSNETS).* Bangalore, India: IEEE.

Bindu, P. V., Mishra, R., & Thilagam, P. S. (2018). Discovering spammer communities in Twitter. *Journal of Intelligent Information Systems, 51*(3), 503–527.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg? *IEEE Transactions On Dependable And Secure Computing, 9*(6), 811-824.

Daffa, W., Bamasag, O., & AlMansour, A. (2018). A Survey On Spam URLs Detection In Twitter. *2018 1st International Conference on Computer Applications & Information Security (ICCAIS).* Riyadh, Saudi Arabia: IEEE.

Fields, J. D. (2016). *Botnet Campaign Detection on* Twitter. Utica, New York: SUNY Polytechnic Institute.

ISSN 2502-3357 (*Online*)
ISSN 2503-0477 (*Print*)

Kenter, T., Borisov, A., & Rijke, M. d. (2016, June 15). *Siamese CBOW: Optimizing Word Embeddings for Sentence Representations.* Retrieved from arXiv:1606.04640: https://arxiv.org/abs/1606.04640

Kuzi, S., Shtok, A., & Kurland, O. (2016). Query Expansion Using Word Embeddings. *CIKM '16 Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (pp. 1929-1932). Indianapolis, Indiana, USA: ACM.

Martinez-Romo, J., & Araujo, L. (2013). Detecting malicious tweets in trending topics using a statistical analysis of language. *Expert Systems with Applications, 40*(8), 2992-3000.

Nguyen, P. T., & Takeda, H. (2016, May 14). *Online learning for Social Spammer Detection on* Twitter. Retrieved from arXiv: https://arxiv.org/abs/1605.04374

Pennington, J., Socher, R., & C. D. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532-1543). Doha, Qatar: Association for Computational Linguistics.

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), 19-25.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*(January), 85-117.

Sedhai, S., & Sun, A. (2018). Semi-Supervised Spam Detection in Twitter Stream. *IEEE Transactions On Computational Social Systems, 5*(1), 169-175.

Yang, C., Harkreader, R. C., & Gu, G. (2011). Die Free or Live Hard? Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers. *International Workshop on Recent Advances in Intrusion Detection* (pp. 318-337). Berlin, Heidelberg: Springer.

Zhang, C. M., & Paxson, V. (2011). Detecting and Analyzing Automated Activity on Twitter. *International Conference on Passive and Active Network Measurement* (pp. 102-111). Berlin, Heidelberg: Springer.