TESIS – EE185401

KLASIFIKASI SPAMMER PADA MICROBLOGGING TWITTER BERDASARKAN PERILAKU PENGGUNA MENGGUNAKAN METODE DECISION TREE DAN NAÏVE BAYES

YULI FITRIANI
07111550050002

DOSEN PEMBIMBING
Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.
Dr. Surya Sumpeno, S.T, M.Eng.

PROGRAM MAGISTER
BIDANG KEAHLIAN JARINGAN CERDAS MULTIMEDIA
DEPARTEMEN TEKNIK ELEKTRO
FAKULTAS TEKNOLOGI ELEKTRO
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2019

TESIS – EE185401

KLASIFIKASI SPAMMER PADA MICROBLOGGING TWITTER BERDASARKAN PERILAKU PENGGUNA MENGGUNAKAN METODE DECISION TREE DAN NAÏVE BAYES

YULI FITRIANI
07111550050002

DOSEN PEMBIMBING
Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.
Dr. Surya Sumpeno, S.T, M.Eng.

PROGRAM MAGISTER
BIDANG KEAHLIAN JARINGAN CERDAS MULTIMEDIA
DEPARTEMEN TEKNIK ELEKTRO
FAKULTAS TEKNOLOGI ELEKTRO
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2019

LEMBAR PENGESAHAN

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Teknik (M.T)
di
Institut Teknologi Sepuluh Nopember

oleh:

Yuli Fitriani
NRP. 07111550050002

Tanggal Ujian : 21 Desember 2018
Periode Wisuda : Maret 2019

Disetujui oleh:

1. Prof. Dr. Ir. Mauridhi Hery P., M.Eng. (Pembimbing I)
   NIP: 19580916 198601 1 001
2. Dr. Surya Sumpeno, S.T., M.Eng (Pembimbing II)
   NIP: 19690613 199702 1 003
3. Dr. Eko Mulyanto Yuniarno, S.T., M.T (Penguji)
   NIP: 19680601 199512 1 009
4. Dr. Diah Puspito Wulandari, S.T., M.Sc (Penguji)
   NIP: 19801219 200501 2 001

Dekan Fakultas Teknologi Elektro

Dr. Tri Arief Sardjono, S.T., M.T.
NIP. 19700212 199512 1 001

Halaman ini sengaja dikosongkan

PERNYATAAN KEASLIAN TESIS

Dengan ini saya menyatakan bahwa isi keseluruhan Tesis saya dengan judul "**KLASIFIKASI SPAMMER PADA MICROBLOGGING TWITTER BERDASARKAN PERILAKU PENGGUNA MENGGUNAKAN METODE DECISION TREE DAN NAÏVE BAYES**" adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diijinkan dan bukan merupakan karya pihak lain yang saya akui sebagai karya sendiri.

Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka. Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Surabaya, Desember 2018

Yuli Fitriani
NRP. 07111550050002

Halaman ini sengaja dikosongkan

KLASIFIKASI SPAMMER PADA MICROBLOGGING TWITTER BERDASARKAN PERILAKU PENGGUNA MENGGUNAKAN METODE DECISION TREE DAN NAÏVE BAYES

Nama mahasiswa : Yuli Fitriani
NRP : 07111550050002
Pembimbing : 1. Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.
2. Dr. Surya Sumpeno, S.T., M.Eng.

ABSTRAK

Twitter merupakan salah satu *Microblogging* yang banyak digunakan oleh masyarakat luas. Popularitas *Twitter* mengundang *spam*mer (penyebar *spam*) untuk mengganggu pengguna lain dengan menyebar *spam tweets* dalam jumlah besar. *Spam tweets* mengandung link URL yang berbahaya, perdagangan obat-obatan terlarang, pornografi, malware, phising. Penelitian ini bertujuan untuk melakukan klasifikasi *spam*mer pada *Microblogging Twitter* berdasarkan perilaku pengguna. Fitur yang digunakan adalah fitur *user-based* dan fitur *content-based*. Penelitian ini menerapkan dua metode klasifikasi, untuk fitur *user-based* menggunakan metode Decision Tree dan untuk fitur *content-based* menggunakan metode Naïve Bayes. Metode klasifikasi Decision Tree dan Naïve Bayes dapat mengklasifikasi *spammer* masing-masing dengan akurasi 88.235%. dan 95.31%.

Kata kunci: decision tree, *microblogging*, naïve bayes, *spam*mer, *twitter*

Halaman ini sengaja dikosongkan

CLASSIFICATION OF *SPAM*MER IN *MICROBLOGGING TWITTER* BASED ON USER'S BEHAVIOR USING DECISION TREE AND NAÏVE BAYES

By : Yuli Fitriani
Student Identity Number : 07111550050002
Supervisor(s) : 1. Prof. Dr. Ir. Mauridhi Hery P. M.Eng.
2. Dr. Surya Sumpeno, S.T., M.Eng.

ABSTRACT

Twitter is one of Microblogging service that widely used by people. Its popularity invites spammers to disturb other users with a large number of spam tweets. Spam tweets contain dangerous URL links, drug trafficking, pornography, malware, phishing. This study aims to classify spammers on Twitter Microblogging based on user behavior. The features used are user-based features and content-based features. This study applies two classification methods, for user-based features using the Decision Tree method and for content-based features using the Naïve Bayes method. The classification method of Decision Tree and Naïve Bayes can classify each spammer with an accuracy of 88.235%. and 95.31%.

Key words: decision tree, *microblogging*, naïve bayes, *spam*mer, *twitter*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Alhamdulillah, segala puji bagi Allah SWT, yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis bias menyelesaikan Tesis yang berjudul "Klasifikasi *Spam*mer pada *Microblogging Twitter* Berdasarkan Perilaku Pengguna Menggunakan Metode Decision Tree dan Naïve Bayes". Naskah tesis ini disusun untuk memenuhi salah satu syarat kelulusan di program studi S2 Teknik Elektro Bidang Keahlian Jaringan Cerdas Multimedia. Pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Allah SWT atas limpahan rahmat-Nya sehingga penulis dapat menyelesaikan Tesis ini dengan baik.
2. Ayah M Takim Santoy (Alm) dan Ibu Hj Suwarti selaku orang tua penulis yang memberi motivasi, doa tiada henti kepada penulis.
3. Prof. Dr. Ir. Mauridhi Hery P., M.Eng dan Dr. Surya Sumpeno, S.T., M.Eng selaku dosen pembimbing yang telah memberi koreksi, doa dan support kepada penulis.
4. Dr. Eko Mulyanto Yuniarno, S.T., M.T selaku koordinator bidang keahlian Jaringan Cerdas Multimedia.
5. Bapak dewan penguji selaku dosen penguji yang telah memberikan saran dan kritik dalam tesis ini.
6. Bapak-bapak dosen pengajar di Program Studi Teknik Elektro, bidang keahlian Jaringan Cerdas Multimedia.
7. Ayah Umar dan Anak-anak tersayang (Shamila Azwa Zuhaira, Syah Kenzo Hidayatullah, dan Shalitta Bintang Khatulistiwa) yang memberi kekuatan jiwa dan raga.
8. Direktur Utama Masjid Nasional Al Akbar Surabaya, Drs. H. Endro Siswantoro, M.Si beserta segenap karyawan/i MAS yang memberi dukungan dan doa kepada penulis.
9. Teman-teman JCM 2015, 2016, 2017 semuanya yang selalu mendoakan dan memberi support kepada penulis dan semua pihak yang telah membantu proses penyelesaian tesis ini.

Sebagai manusia biasa, penulis menyadari bahwa Tesis ini masih jauh dari kata sempurna dan memiliki banyak kekurangan. Sehingga dengan segala kerendahan hati, penulis mengharapkan saran dan kritik yang membangun dari pembaca.

Surabaya, 1 Desember 2018
Penulis

DAFTAR ISI

DAFTAR GAMBAR

Halaman ini sengaja dikosongkan

DAFTAR TABEL

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Situs media sosial seperti *Facebook, MySpace, dan Twitter* memberikan peluang besar kepada pengguna untuk berkenalan dengan pengguna lain, memperkenalkan produk barang/jasa yang dimiliki, memperluas koneksi, memperlebar jangkauan dalam profesionalitas kerja, memberikan informasi hasil karya terbaru, mengirim berita terkini dll. Masyarakat dapat berbagi dan memperoleh informasi dengan mudah melalui situs-situs media sosial tersebut. *Twitter* adalah salah satu *Microblogging* yang banyak digunakan oleh masyarakat luas [1]. *Twitter term*asuk dalam 15 besar daftar situs media sosial yang paling populer. *Twitter* memberi fasilitas kepada penggunanya untuk mengirim text, yang disebut *tweets* dengan batas 140 karakter. Selain itu pengguna juga dapat mengirim video singkat, link URL, dan juga gambar.

Di sisi lain, popularitas *Twitter* juga memberi kesempatan bagi penyebar *spam (spammer)* untuk menyebarkan *spam*. *Spam tweets* mengandung link URL yang berbahaya, perdagangan obat-obatan terlarang, pornografi, *malware*, *phishing* [2]. Untuk mengatasi penyebaran *spam*, *Twitter* memberi solusi untuk semua penggunanya dengan cara melaporkan akun yang ditengarai sebagai *spam*. Pengguna dapat meng-klik "*report tweet*" pada *homepage* akun *spam*. Dengan demikian, pihak *Twitter* akan melakukan penelusuran perihal akun yang dilaporkan tersebut. Jika memang benar bahwa akun tersebut adalah akun penyebar *spam*, maka *Twitter* akan men-suspend akun tersebut sehingga pengguna lain tidak dirugikan karenanya [3]. *Twitter* mempunyai banyak peraturan untuk dipahami oleh semua penggunanya, salah satunya adalah dengan tidak menyebarkan *spam* [4].

Jika Anda memiliki jumlah "*follow*ing" lebih banyak dari pada jumlah "*follow*ers", maka Anda dapat ditengarai sebagai akun *spam* [5]. Kebiasaan *spam*mer adalah dengan melakukan "*follow*" kepada banyak akun dalam waktu yang singkat. Sehingga jumlah "*follow*ing" akun tersebut menjadi lebih banyak dibandingkan jumlah "*follow*ers" nya. Perilaku *spammer* dengan pengguna normal

sangat berbeda. *Spammer* selalu mengirimkan *spam tweets* secara berulang dalam waktu yang singkat kepada sebanyak-banyaknya pengguna lain. Spammer juga banyak mempergunakan *trending topic* (topik populer) pada *Twitter* untuk menyebarkan *spam tweets* [6]. Hal ini "memaksa" pengguna lain untuk membaca *posting*an *spammer*, padahal sebenarnya pengguna yang lain bermaksud untuk membaca perkembangan dari *trending topic* itu sendiri. Namun yang didapat adalah tumpukan tweet sampah.

Keberadaan *spam* yang mengganggu pengguna lain menjadi perhatian dalam penelitian ini. Untuk itu dibutuhkan suatu sistem dalam mengklasifikasikan bahwa suatu akun merupakan akun *spam*mer dan non-*spam*mer. Penelitian ini mengumpulkan data akun-akun pada *Twitter* dengan menggunakan API (*Application Programming Interface*) *Twitter*. Klasifikasi ini menjadi penting dengan maraknya penyebaran *spam* yang tidak hanya menyerang akun pengguna biasa, namun juga merambah ranah akun pengguna authentic (*verified account*). Dengan meneliti perilaku *spam*mer yang "tidak biasa" akan memberikan informasi bahwa akun ini merupakan akun *spam*mer atau akun biasa.

Untuk membahas perilaku *spam*mer pada *microblogging Twitter* tidak cukup hanya dilihat dari konten/isi dari pengguna, yang secara kasat mata, yang dapat dibaca oleh siapapun, namun juga perlu untuk dikaji profil dari pengguna itu sendiri. Oleh karena itu, penelitian ini menggunakan fitur *user-based* dan fitur *content-based*. Semua fitur user-based dan fitur *content-based* adalah termasuk perilaku *spammer* pada *Twitter* yang dibahas pada penelitian ini.

Kategori fitur *user-based* adalah hubungan pengguna seperti orang- orang yang pengguna ikuti dan orang-orang yang mengikuti pengguna atau perilaku pengguna seperti frekuensi pengguna mengirim *tweet*, limit selang waktu antara *tweet* pertama dan seterusnya, apakah pengguna adalah termasuk pengguna yang terverifikasi oleh *Twitter* atau tidak. Sedangkan kategori fitur *content-based* adalah kata-kata dari konten yang dikirim oleh pengguna itu sendiri. Tentunya kata-kata tersebut telah melewati tahap praproses terlebih dahulu sehingga didapatkan kata-kata yang termasuk *term* (token unik).

Ada berbagai macam algoritma untuk mengklasifikasikan *spammer*. Dalam penelitian ini, untuk fitur *user-based* menggunakan metode Decicion Tree,

sedangkan untuk fitur *content-based* menggunakan metode Naïve Bayes. Karena adanya dua dataset dari masing-masing fitur *user-based* dan fitur *content-based* yang strukturnya berbeda, banyak dokumen dan fitur-fitur yang digunakan juga berbeda, maka digunakan dua metode klasifikasi yang telah disebutkan sebelumnya. Decision Tree diterapkan pada fitur *user-based* karena dataset yang didapatkan adalah merupakan data statistik yang implementasinya lebih mudah menggunakan model pohon keputusan. Decision Tree merupakan algoritma pengklasifikasian yang sering digunakan dan mempunyai struktur yang sederhana dan mudah untuk diinterpretasikan [7]. Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk membuat proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan.

Sedangkan untuk *content-based* diterapkan metode Naïve Bayes karena dataset yang telah melalui tahap praproses berbentuk matriks yang disebut *term*-document matrix dengan dimensi yang lebar, sehingga jika diterapkan metode Decision Tree akan tidak optimal hasilnya, cabang (*branches*) atau node yang tidak diperlukan dapat menyebabkan ukuran Decision Tree menjadi sangat besar dan hal ini disebut *over-fitting* [7]. Untuk itu digunakan metode Naïve Bayes yang berfungsi menyederhanakan data untuk klasifikasi, selain itu Naïve Bayes banyak digunakan dalam pengklasifikasian dokumen karena kesederhanaan algoritma dan mudah untuk diimplementasikan [8].

1.2 Rumusan Masalah

Aktivitas *spam*ming tidak berhenti pada sekedar mengirimkan *tweet* sampah, tetapi telah menjurus ke arah penipuan dan hal-hal merugikan lainnya. *Spam* merupakan penyalahgunaan dalam pengiriman berita dari jaringan komunikasi dan memiliki berbagai bentuk dan definisi yang berbeda tergantung pada jenis jaringannya. Akibatnya banyak pengguna yang merasa terganggu oleh banyaknya *tweet* sampah tersebut, untuk mengatasi hal ini diperlukan suatu filter yang dapat membedakan *spam*mer atau non*spam*mer.

1.3 Tujuan

Penelitian bertujuan untuk mengidentifikasi dan mengklasifikasi *spam*mer dan non-*spam*mer dengan mengetahui karakteristik perilaku pengguna (dari segi *user-based* dan *content-based)* sehingga dapat membantu masyarakat dalam berhati-hati dan bijak memanfaatkan layanan jejaring sosial *Twitter*.

1.4 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah :

1. Penelitian ini dilakukan pada media sosial *Twitter*, dengan jumlah 100 akun *Twitter* dengan pengambilan (*crawling*) 18-25 tweets terakhir masing-masing akun.
2. Penelitian ini mempelajari klasifikasi *spam*mer dan non-*spam*mer berdasarkan perilaku pengguna *Twitter* dengan fitur *user-based* menggunakan metode Decision Tree, sedangkan fitur *content-based* menggunakan metode Naïve Bayes.
3. Studi kasus di Indonesia.
4. Data akun *Twitter* diambil sendiri oleh penulis dengan menggunakan API *Twitter*.

1.5 Kontribusi

Kontribusi dari penelitian ini adalah membantu organisasi, *term*asuk organisasi pemerintah dan jasa layanan darurat dan seluruh lapisan masyarakat untuk merespon postingan *Twitter* lebih efektif dan bijaksana. Selain itu, masyarakat dapat menemukan telaah antara informasi yang benar dan informasi yang berupa *spam* pada data *tweet* di masyarakat pada microblog *Twitter*.

BAB 2
KAJIAN PUSTAKA

2.1 Kajian Penelitian Terkait

Permasalahan *spam* yang menyebar pada *Microblogging Twitter* telah sebelumnya dipelajari oleh peneliti. Beberapa peneliti mempelajari karakteristik dari *spam* itu sendiri. Kemudian dilakukan beberapa percobaan dalam mendeteksi *spam* dengan beberapa metode yang berbeda.

Seperti pada penelitian [5] yang berupaya mengklasifikasi *spam* pada *Twitter*. Mereka menggunakan beberapa fitur pada masing-masing akun *Twitter*. Antara lain fitur pada akun *Twitter* dan fitur pada konten /pesan yang dikirim. Metode yang digunakan adalah Random Forest, hasil yang didapatkan adalah 95,7% dapat memprediksi label dengan benar.

Penulis pada [9] menggunakan metode Naïve Bayes untuk mendekteksi akun *spam* pada *Twitter*. Mereka meneliti *spam tweet*s dan non-*spam tweet*s. Rasio deteksi *spam tweet*s yang dapat diklasifikasi dengan benar adalah sebesar 75,3%.

Twitter menjadi populer dengan jumlah pengguna yang bertambah secara signifikan dari tahun ke tahun, menyebabkan persebaran *spam* semakin mudah. Dan semakin mudah juga untuk mendeteksi karakteristik dari *spam*mer itu sendiri. Pada [6] menyebutkan bahwa ada tujuh karakteristik *spam*mer. Karakteristik yang paling menonjol adalah *spam*mer mengirimkan pesan yang sama dalam ke banyak pengguna dalam waktu yang singkat. *Spam*mer mempunyai beberapa akun duplikat, hal ini dikarenakan mereka banyak dilaporkan oleh pengguna lain sebagai *spam* kepada *Twitter*, selain itu mereka juga sering di-block oleh pengguna lainnya karena pesan yang mengganggu. Namun, beberapa *spam*mer biasanya menyebut pengguna lain (*mention*) dalam setiap pesannya dengan menambahkan "@namapenggunalain" untuk menarik perhatian pengguna tersebut. Sehingga pengguna yang disebutkan itu, mau tidak mau, langsung atau secara tidak langsung, membaca pesan yang disebarkan *spam*mer.

Tujuan utama dari *spam*mer adalah untuk mencari uang. Untuk mencari uang lebih banyak lagi, *spam*mer akan mengajak banyak pengguna untuk membaca

pesannya dan meng klik link URL yang ditautkan [10] dan dapat pula link tersebut membahayakan pengguna lain [6]. Pada penelitian ini [6] digunakan metode Random Forest untuk klasifikasi *spam*mer dan non-*spam*mer. Rasio *spam*mer yang dapat diklasifikasi dengan benar adalah sebesar 75,0%.

Penulis pada [5] menggunakan metode Support Vector Machines (SVM), Random Forest, K-Nearest Neighbor (KNN), dan Naïve Bayes untuk mendeteksi *spam*mer pada *Twitter*. Data dikumpulkan menggunakan *Twitter* API [11]. Hasil terbaik adalah menggunakan metode Random Forest.

Pada penelitian [12] menggunakan Threshold dan Associative dan Support Vector Machines (SVM) untuk mengklasifikasikan profil akun *spam*. Hasil dari metode Threshold dan Associative didapatkan akurasi sebesar 79,26%. Untuk metode SVM akurasi yang didapatkan sebesar 69,32%.

Teknik Hibrid dilakukan pada penelitian [13]. Mereka menggunakan fitur pada profil akun pengguna dan fitur pada konten /pesan yang dikirim untuk mengidentifikasi *spam*mer pada *Twitter*. Dengan beberapa metode yang digunakan, yakni J48, Decorate, dan Naïve Bayes, hasil terbaik adalah metode J48 dan Decorate dengan presisi yang dicapai sebesar 97,6%.

2.2 *Microblogging Twitter*

Microblogging adalah salah satu jenis komunikasi, pengguna dapat menulis dan menyebarkan teks yang dikirim melalui pesan teks singkat, email atau web [14]. Jenis-jenis microblog antara lain *Facebook, MySpace*, *Twitter*. *Twitter* adalah salah satu layanan microblog yang banyak digunakan oleh masyarakat luas [1]. *Twitter* menyediakan layanan untuk mengirim teks dengan batasan 140 karakter [15]. Selain itu, *Twitter* juga dapat digunakan untuk mengirim video, link URL, dan juga gambar.

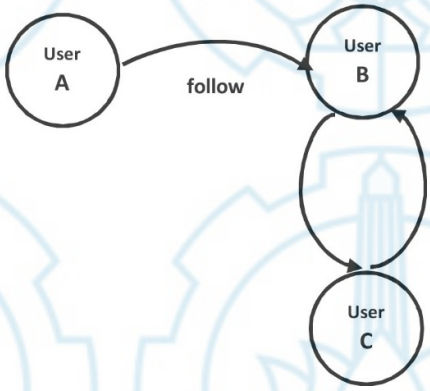
Twitter mengalami perkembangan yang sangat pesat. Selain dapat diakses melalui PC (personal computer), *Twitter* juga dapat diakses melalui perangkat lunak seluler, dan juga aplikasi *Twitter*. Pengguna dapat registrasi untuk dapat menggunakan layanan yang disediakan oleh *Twitter*. Pengguna dapat dikenali melalui *user*name atau nama sebenarnya. Pada *Twitter* pengguna tidak hanya bisa mengirim dan membaca *tweet* tetapi pengguna juga bisa memberi balasan,

meneruskan pesan, re*tweet*, *mention* pengguna lain, menggunakan *hashtag* (#) untuk topik tertentu, dan juga melakukan *follow*ing atau mem-*follow* kembali.

Pengguna A dapat melakukan "*follow*ing" pada pengguna B. Pengguna B yang di"*follow*" dapat mem"*follow*" kembali pengguna A. Mem"*follow*" atau tidak mem"*follow*" suatu akun adalah hak setiap pengguna. Jika pengguna B tidak suka dengan pengguna A, maka pengguna B bisa mengacuhkan hal itu. Pengguna B mem"*follow*" pengguna C dan pengguna C mem"*follow*" pengguna B. Pengguna B dan pengguna C adalah teman (friends) dalam istilah *Twitter*.

Pengguna A adalah "*follow*er" pengguna B. Sebagai *follow*er, maka pengguna A akan menerima setiap pesan (text, gambar, video) yang dikirimkan oleh pengguna B. Setiap kali pengguna B mengirim pesan, maka seluruh *follow*ernya akan menerima pesan tersebut. Pengguna B dapat melakukan proteksi terhadap akunnya, dalam hal ini pengguna lain dapat mem*follow* akun pengguna B jika diijinkan oleh pengguna B. Ini adalah salah satu cara bentuk kehati-hatian dalam melakukan seleksi untuk *follow*ers yang berdatangan.

Semua pesan pada *Twitter* dapat disatukan dengan sebuah *hashtag* (#) untuk suatu topik tertentu. Misalnya #prayforPalu, hal ini menunjukkan bahwa isi pesan yang berhubungan dengan Palu, *term*asuk doa dari masyarakat Indonesia, donasi untuk Palu, dan juga apapun yang berhubungan dengan gempa yang melanda Palu, Indonesia, tergabung dengan adanya *hashtag* (#) tersebut, sehingga mempermudah pengguna lain dalam mencari informasi seputar gempa Palu. Topik yang ditandai dengan *hashtag* dinamai trending topics. Grafik *Twitter* secara sederhana dapat dilihat pada Gambar 2. 1.



Gambar 2. 1 Grafik *Twitter* sederhana [3]

2.3 *Twitter* API

API (Application Programming Interface) merupakan sebuah aplikasi yang diciptakan *Twitter* untuk developer mengakses informasi [16]. *Twitter* API terdiri dari dua komponen yang berbeda, REST dan SEARCH API. REST API memungkinkan developer *Twitter* untuk mengakses data core *Twitter* (*tweet*, timeline, *user* data). SEARCH API digunakan untuk membuat query *tweet*, *term*asuk menyediakan informasi tentang trending topics.

Layanan API memperbolehkan para developer (pengembang) aplikasi mobile untuk mengintegrasikan aplikasi mereka dengan *Twitter*. Oleh karena itu, banyak aplikasi penyedia layanan berita menggabungkan konten berita dari banyak media sosial menggunakan API.

2.4 *Spam* (*Stupid Pointless Annoying Message*)

Semua orang yang berhubungan dengan internet dan sosial media pasti tidak asing lagi dengan kata *spam* (*Stupid Pointless Annoying Message*). *Spam* adalah pesan atau email yang dikirimkan secara massal tanpa dikehendaki oleh penerimanya [10]. *Spam*mer adalah seseorang yang berusaha mengirimkan pesan massal kepada sebanyak-banyaknya pengguna untuk mempromosikan suatu produk/jasa atau merusak sistem pengguna lain. Tindakan menyebarkan *spam* disebut dengan *spam*ming, sedangkan orang yang melakukan *spam* disebut *spam*mer.

2.5 Deteksi *Spam*mer

Twitter menjadi populer dengan jumlah pengguna yang bertambah secara signifikan dari tahun ke tahun, menyebabkan persebaran *spam* semakin mudah. Dan semakin mudah juga untuk mendeteksi karakteristik dari *spam*mer itu sendiri. Pada [6] menyebutkan bahwa ada tujuh karakteristik *spam*mer. Karakteristik yang paling menonjol adalah *spam*mer mengirimkan pesan yang sama dalam ke lebih banyak pengguna dalam waktu yang singkat.

*Spam*mer mempunyai beberapa akun duplikat, hal ini dikarenakan mereka banyak dilaporkan oleh pengguna lain sebagai *spam* kepada *Twitter*, selain itu

mereka juga sering di-block oleh pengguna lainnya karena pesan yang mengganggu. Namun, beberapa *spam*mer biasanya menyebut pengguna lain (*mention*) dalam setiap pesannya dengan menambahkan "@namapenggunalain" untuk menarik perhatian pengguna tersebut. Sehingga pengguna yang disebutkan itu, mau tidak mau, langsung atau secara tidak langsung, membaca pesan yang disebarkan *spam*mer.

Gambar 2. 2 Contoh halaman *spam* pada *Twitter*

Selain itu, *spam*mer mempunyai beberapa teknik untuk mem *follow* pengguna lain dalam waktu yang singkat. Dengan cara membeli *follow*ers, saling mem *follow* akun masing-masing untuk memperbesar jumlah *follow*ers dan *follow*ing mereka [10]. Seorang *spam*mer juga tidak perlu ada hubungan apapun dengan korbannya, baik itu mem *follow* atau menjadi *follow*er. Cara lain adalah

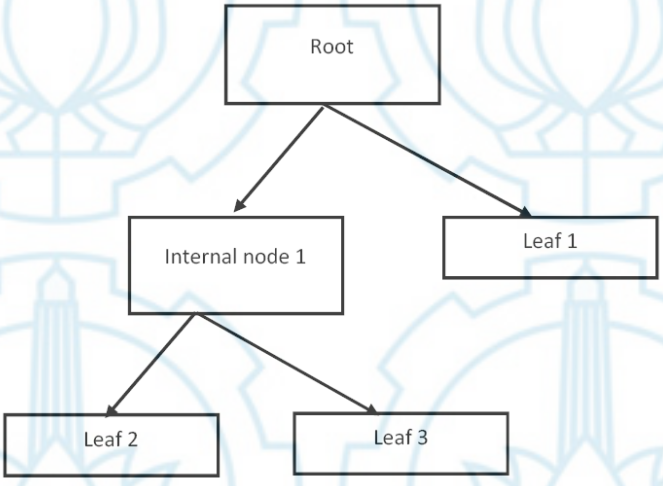
berpura-pura menggunakan *trending topics* dengan cara menambahkan *hashtag* pada setiap pesannya seolah-olah pesan yang dikirim adalah bagian dari trending topics tersebut. Kenyataannya adalah pesan *spam* yang bertumpuk. Contoh halaman *spam* pada *Twitter* dapat dilihat pada Gambar 2.2.

Pada penelitian [6] disebutkan beberapa karakteristik *spam* secara keseluruhan :

1. Mengirim link URL yang berbahaya
2. Perilaku yang agresif
3. Menyalahgunakan @reply atau @*mention*
4. Membuat banyak akun
5. Mengirim pesan berulang kali tentang topik yang sedang tren (trending topics)
6. Mengirim pesan pembaruan yang sama berulang kali
7. Mengirim link URL dengan isi pesan yang tidak terkait

2.6 Decision Tree

Pohon keputusan (Decision Tree) adalah algoritma klasifikasi yang mempunyai struktur sederhana dan mudah untuk aplikasinya [7]. Decision Tree merupakan model klasifikasi yang berbentuk pohon terbalik, dimana akar (root) berada di bagian paling atas, dan daun (leaf) berada di bagian paling bawah. Model klasidikasi ini mudah dipahami dan lebih efisien dalam menginduksi data.



Gambar 2. 3 Grafik sederhana Decision Tree

Metode Decision Tree ini cocok digunakan untuk klasifikasi atau prediksi [17]. Grafik sederhana Decision Tree dapat dilihat pada Gambar 2.3.

Proses pada Decision Tree adalah mengubah bentuk data (tabel) menjadi pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule [18]. Konsep Decision Tree secara garis besar dapat dilihat pada Gambar 2.4

Gambar 2. 4 Konsep Decision Tree

Manfaat utama dari penggunaan Decision Tree adalah kemampuannya dalam mengambil suatu keputusan yang kompleks menjadi lebih sederhana sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari suatu permasalahan.

2.7 Naïve Bayes

Selain menggunakan Decision Tree untuk klasifikasi apakah suatu akun merupakan akun *spam*mer atau bukan, penelitian ini juga meneliti konten/isi dari pesan *spam* itu sendiri. Dengan menggunakan Naïve Bayes, data pesan dari akun yang telah dikumpulkan tersebut akan diklasifikasikan menjadi dua kelas yaitu *spam*mer dan non-*spam*mer.

Naïve Bayes adalah salah satu algoritma yang digunakan untuk klasifikasi teks serta merupakan metode *Machine Learning* yang menggunakan perhitungan probabilitas dan statistik yang dikemukakan oleh Thomas Bayes. Algoritma tersebut digunakan untuk memprediksi probabilitas di masa depan berdasarkan pengalaman di masa lalu.

Naïve Bayes merupakan salah satu algoritma klasifikasi yang menggunakan statistika. Pengklasifikasian ini menggunakan probabilitas sederhana yang diadopsi dari teorema Bayes (statistik Bayesian), dimana menggunakan asumsi naif atau independen yang kuat. Dalam proses pengklasifikasian teks terdapat dua tahapan, yaitu: tahap pelatihan (training) dan tahap pengujian (testing). Tahap pelatihan

merupakan tahap pelatihan terhadap sejumlah dokumen contoh, sedangkan tahap pengujian merupakan proses pengklasifikasian dokumen baru dan belum diketahui kategorinya. Dalam *term*inologi sederhana, sebuah Naïve Bayes Classifier mengasumsikan bahwa kehadiran (atau ketiadaan) fitur tertentu dari suatu kelas tidak berhubungan dengan kehadiran (atau ketiadaan) fitur lainnya. Sebagai contoh, buah mungkin dianggap apel jika merah, bulat, dan berdiameter sekitar 4 inchi. Bahkan jika fitur ini bergantung satu sama lain atau atas keberadaan fitur lain. Sebuah NBC menganggap bahwa seluruh sifat-sifat berkontribusi mandiri untuk probabilitas bahwa buah ini adalah apel.

Sebuah keuntungan dari Naïve Bayes Classifier adalah bahwa Naïve Bayes Classifier memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata dan varian dari variabel) yang diperlukan untuk klasifikasi. Karena variabel diasumsikan independen, hanya varian dari variabel-variabel untuk setiap kelas yang perlu ditentukan dan bukan keseluruhan covariance matrix. Model probabilitas untuk classifier adalah model kondisional pada persamaan 2.1.

$$p(C|F1, \ldots., Fn) \tag{2.1}$$

terhadap variabel kelas dependen C dengan sejumlah kecil hasil atau kelas, tergantung pada beberapa variabel fitur F1 sampai Fn. Perlu dicermati bahwa jika jumlah fitur n besar atau bila fitur bisa mengambil sejumlah besar nilai, maka membuat sebuah model pada tabel probabilitas adalah tidak mungkin. Oleh karena itu diperlukan reformulasi model untuk membuatnya lebih fleksibel. Menggunakan teorema Bayes seperti pada persamaan 2.2.

$$p(C|F1, \ldots. Fn) = \frac{p(C)p(F1,\ldots.,Fn|C)}{p(F1,\ldots,Fn)} \tag{2.2}$$

2.8 Pembobotan

Pada penelitian ini, konten yang terdiri dari kata-kata yang telah melalui praproses akan menjadi sebuah *term*. *Term* adalah token unik. *Term* yang telah

terbentuk dihitung bobot kemunculannya dengan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF tersebut dilakukan untuk melihat bobot keterkaitan suatu *term* dengan dokumen. *Term Frequency* (TF) merupakan banyaknya *term* yang muncul pada suatu dokumen. Sedangkan *Inverse Document Frequency* (IDF) bertujuan untuk mengetahui apa saja kata kunci yang menyusun dokumen tersebut. *Term* yang sering muncul akan memberikan pengaruh yang kecil dalam menentukan keterkaitan kata kunci dengan dokumen. TF-IDF dapat dihitung dengan persamaan 2.3.

$$w_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \cdot log_2 \frac{D}{d_i} \quad (2.3)$$

2.9 Perhitungan Kinerja

Penelitian ini menggunakan matriks standar untuk mengevaluasi hasil dari metode yang telah digunakan. Tabel 2.1 menunjukkan confusion matrix untuk klasifikasi *spam*mer.

Tabel 2. 1. Confusion Matrix

		Prediction	
		Spammer	Non-spammer
True	Spammer	a	b
	Non-spammer	c	d

Dimana a menggambarkan akun dari kelas *spam*mer yang diklasifikasikan secara benar sebagai *spam*mer, b adalah akun dari *spam*mer yang seharusnya diklasifikasikan sebagai *spam*mer, namun pada proses klasifikasi, b terklasifikasi menjadi kelas non-*spam*mer, c menunjukkan kelas non-*spam*mer yang setelah proses klasifikasi, c terklasifikasi menjadi kelas *spam*mer, sedangkan d adalah kelas non-*spam*mer yang diklasifikasikan secara benar sebagai kelas non-*spam*mer.

Berdasarkan keempat definisi tersebut, diperoleh nilai Accuracy, Precision, Sensitivity, Specificity. Masing-masing dapat dijelaskan sebagai berikut:

1. *Accuracy* adalah prosentase keakuratan metode klasifikasi yang digunakan untuk melakukan klasifikasi terhadap *spam*mer dan non-*spam*mer. Perhitungan *Accuracy* diperoleh dari persamaan 2.4

$$Accuracy = \frac{a+d}{a+d+c+b} \quad (2.4)$$

2. *Precision* artinya tingkat kebenaran metode ini dalam melakukan klasifikasi dengan benar. Perhitungan *Precision* diperoleh dari persamaan 2.5

$$Precision = \frac{a}{a+c} \quad (2.5)$$

3. *Sensitivity* artinya prosentase akun *spam*mer yang dapat diprediksi dengan benar. Perhitungan *sensitivity* diperoleh dari persamaan 2.6

$$Sensitivity = \frac{a}{a+b} \quad (2.6)$$

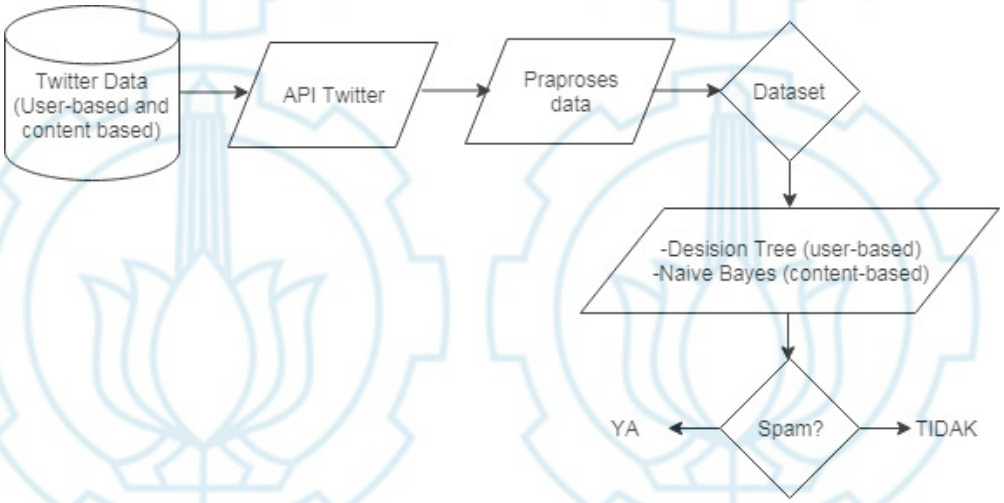
4. *Specificity* adalah prosentase akun non-*spam*mer yang dapat diprediksi secara benar. Perhitungan *specivicity* diperoleh dari persamaan 2.7

$$Specificity = \frac{d}{d+c} \quad (2.7)$$

BAB 3
METODOLOGI PENELITIAN

3.1 Data Penelitian

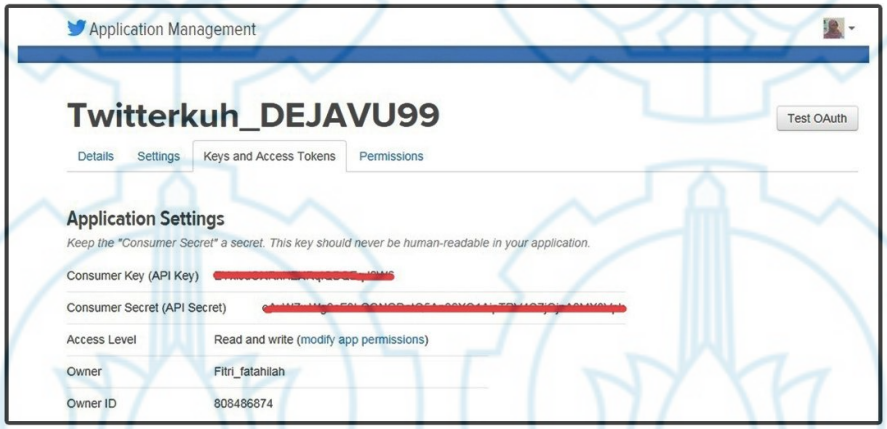
Data pada penelitian ini adalah hasil dari *crawling* (pengumpulan) data melalui API *Twitter*. Data ini terdiri dari profil akun pengguna *Twitter* yang *term*asuk *spam*mer dan non-*spam*mer yang berjumlah 100 akun. Akun *spam*mer berjumlah 51 akun, sedangkan akun non-*spam*mer berjumlah 49 akun. Akun non-*spam*mer terdiri dari 25 akun pengguna biasa dan 24 akun lembaga/instansi. Dari 100 akun tersebut, juga diambil 18 sampai 25 *tweet*s terakhir dari masing-masing akun. Sehingga didapatkan 2156 *tweet* dari akun *spam*mer dan non-*spam*mer. Blok diagram sistem dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Blok diagram sistem

Sebelum proses crawling data melalui *Twitter* API, langkah pertama yang harus dilakukan adalah membuat koneksi dengan *Twitter* API, yang sebelumnya pengguna harus mempunyai akun di *Microblogging Twitter* terlebih dahulu. Selanjutnya adalah membuat aplikasi manajemen pada *Twitter* dengan klik laman

https://apps.*Twitter*.com. Contoh aplikasi manajemen pada *Twitter* dapat dilihat pada Gambar 3.2.



Gambar 3. 2 Contoh aplikasi manajemen pada *Twitter*

Setelah berhasil membuat aplikasi manajemen pada *Twitter*, maka *Twitter* akan mengirim consumer key (API *key*) dan consumer secret (API *secret*). API key dan API secret untuk setiap pembuatan aplikasi manajemen pada *Twitter* adalah berbeda-beda, dan ini sifatnya unik, artinya setiap pengguna yang menggunakan fasilitas ini akan mendapatkan API *key* dan API *secret* yang berbeda-beda. API *key* dan API *secret* ini yang kemudian digunakan untuk melakukan otorisasi pada *Twitter*.

Langkah selanjutnya adalah membuat koneksi *Rstudio* [19] dengan *Twitter* API. *Rstudio* membutuhkan package *Twitter* untuk tersambung dengan *Twitter* API. Package *Twitter* membutuhkan tiga package yaitu Rcurl, ROAuth dan rjson. Package ROAuth digunakan untuk proses otorisasi. Proses otorisasi membutuhkan API key dan API secret. Pada proses otorisasi, R akan secara otomatis membuka browser untuk melakukan verifikasi pada aplikasi manajemen *Twitter* yang telah dibuat dan akan muncul kode numerik yang acak pada browser. Akhirnya, proses verifikasi telah selesai.

Tahapan dilanjutkan dengan mengumpulkan (*crawling*) data pada *Twitter*. Dalam hal ini adalah akun–akun pengguna *Twitter* yang *term*asuk *spam*mer maupun non-*spam*mer. Maka terkumpullah 100 akun. Akun *spam*mer berjumlah 51

akun, sedangkan akun non-*spam*mer berjumlah 49 akun. Akun non-*spam*mer terdiri dari 25 akun pengguna biasa dan 24 akun lembaga/instansi/akun ter-verifikasi *Twitter*, seperti @detikcom, @aagym, dll.

Untuk memberi label bahwa akun ini *term*asuk *spam*mer atau bukan, maka dapat dilihat dari pesan yang dikirimkan masing-masing akun. Dapat dilihat pada kajian sebelumnya untuk mendeteksi *spam*mer adalah kembali pada karakteristik dari *spam*mer itu sendiri. Pada [6] menyebutkan bahwa ada tujuh karakteristik *spam*mer. Karakteristik yang paling menonjol adalah *spam*mer mengirimkan pesan yang sama dalam ke lebih banyak pengguna dalam waktu yang singkat.

Setelah semua langkah terpenuhi, maka dapat dilihat pada dataset yang telah terkumpul, bahwa pada akhirnya setiap akun akan mempunyai tujuh atribut yaitu jumlah *tweet*, jumlah *follow*er, jumlah *follow*ing, usia akun, rataan *tweet* per hari, rataan selang waktu antar *tweet*, dan *verified user*.

3.2 Desain Sistem Decision Tree

Pada tahap ini, metode yang digunakan dalam pengklasifikasian *spam*mer berdasarkan perilaku pengguna (*user-based features*) adalah Decision Tree. Secara umum proses ini dibagi menjadi beberapa tahap. Tahap-tahap tersebut dapat dilihat pada Gambar 3.3.

Tahap-tahap pengklasifikasian *spam*mer menggunakan Decision Tree adalah sebagai berikut :

1. Menghitung *entropy*

   Entropy dapat diperoleh dengan persamaan 3. 1.

$$Entropi\ (S) = \sum_{j=1}^{k} - p_j\ \log_2\ p_j \tag{3.1}$$

   Keterangan :

   • *S* adalah himpunan (dataset) kasus

   • *k* adalah banyaknya partisi *S*

   • p_j adalah probabilitas yang di dapat dari Sum(Ya) dibagi Total Kasus

2. Menghitung nilai *gain ratio* masing-masing atribut

Nilai *gain ratio* dapat dihitung dengan persamaan 3. 2, persamaan 3. 3, dan persamaan 3. 4

$$gain\ ratio(a) = \frac{gain(a)}{split(a)} \tag{3.2}$$

Dimana:

a = atribut.

gain(a) = *information gain* pada atribut a

Split(a) = *split information* pada atribut a

$$SplitInfo(S, A) = -\sum_{i=1}^{n} \frac{S_i}{S} \log_2 \frac{S_i}{S} \tag{3.3}$$

Dimana:

S = ruang (data) *sample* yang digunakan untuk training.

A = atribut.

S_i = jumlah *sample* untuk atribut i

$$Gain\ (A) = Entropi\ (S) - \sum_{i=1}^{k} \frac{|S_i|}{|S|} \times Entropi(S_i) \tag{3.4}$$

Dimana:

S = ruang (data) *sample* yang digunakan untuk training.

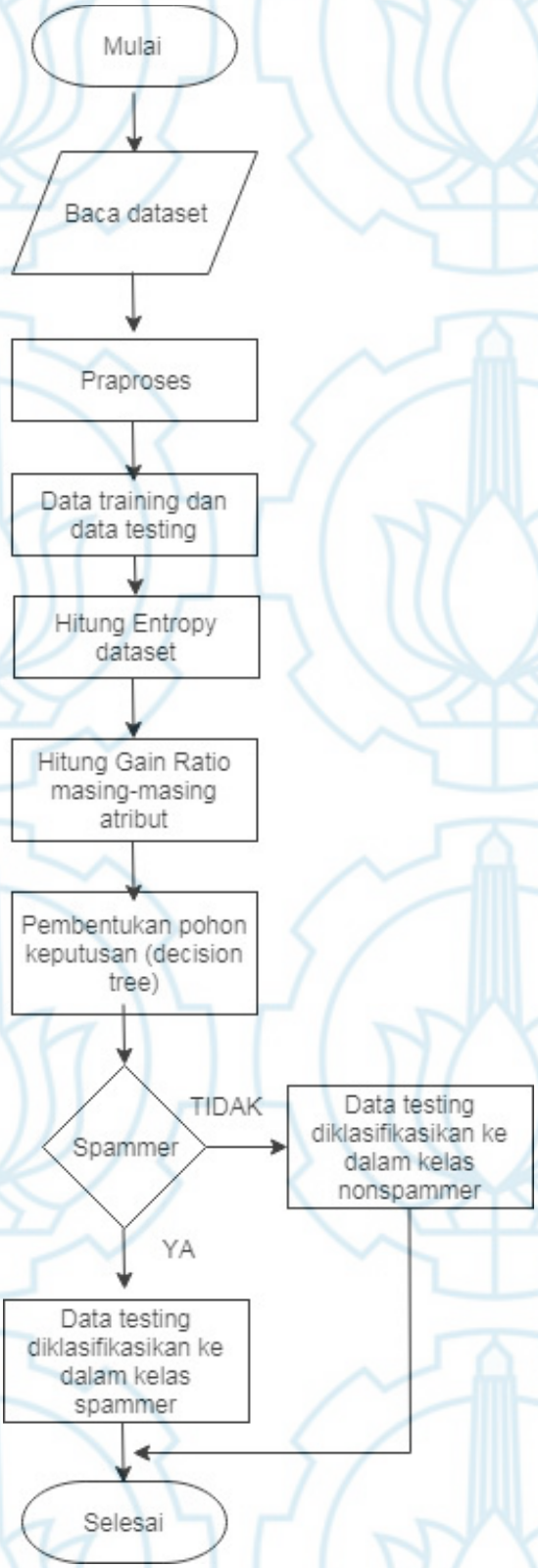
A = atribut.

$|S_i|$ = jumlah sample untuk nilai V.

|S| = jumlah seluruh *sample* data.

Entropi(S_i) = *entropy* untuk sample-sample yang memiliki nilai *i*

3. Atribut dengan gain ratio paling besar menjadi akar, dan atribut lebih rendah menjadi cabang
4. Menghitung lagi nilai gain ratio tiap atribut dg tidak mengikutsertakan lagi atribut yg telah menjadi akar pd tahap sebelumnya
5. Atribut yg tertinggi menjadi cabang
6. Ulangi langkah 4 dan 5 sampai nilai Gain=0 untuk semua atribut yg tersisa



Gambar 3. 3 Desain sistem Decision Tree

3.3 Praproses Data *Spam*mer Berdasarkan Profil Pengguna (*User*-based Features)

Sebelum data siap untuk diolah, maka terlebih dahulu harus dilakukan praproses data. Tujuan dari praproses data adalah untuk mentransformasi data ke suatu format yang prosesnya menjadi lebih mudah untuk gunakan pada langkah selanjtnya. Pada penelitian ini ada dua tahap untuk proses praproses data, yaitu tahap seleksi data dan tahap transformasi data :

1. Seleksi Data

Pada tahap ini digunakan fungsi get*User* dan *user*Timeline pada package *Twitter*. Fungsi get*User* digunakan untuk memilih atribut data yang akan diklasifikasi. Sedangkan fungsi *user*Timeline adalah memilih atribut pada transformasi data.

Diantara atribut data yang dipilih pada fungsi get*User* adalah jumlah *tweet*, jumlah *follow*er, jumlah *follow*ing, usia akun. Pemilihan atribut pada fungsi get*User* dapat dilihat pada Tabel 3.1.

Tabel 3. 1 Pemilihan atribut data pada fungsi get*User*

Function	Attributes	Note
getUser	followersCount	Number of user's followers
	friendsCount	Number of user's following
	created	The date when account has been created
	statusesCount	Number of user's statuses

2. Transformasi Data

Pada tahap ini dilakukan pemilihan atribut pada data yang telah terkumpul dan juga menggabungkan atribut yang ada pad fungsi get*User* dan fungsi *user*Timeline. Pada akhirnya tiap-tiap data akan mempunyai atribut jumlah *tweet*, jumlah *follow*er, jumlah *follow*ing, usia akun, rataan *tweet* per hari, rataan selang waktu antar *tweet*, dan *verified user*. Contoh dataset kecil pada fungsi *user*Timeline dapat dilihat pada Tabel 3.2.

Tabel 3. 2 Contoh dataset kecil pada fungsi *user*Timeline

	text	favorited	favoriteCount	replyToSN	created	truncated	replyToSID
1	Hai Ka @adistidistiadis Suka Masak? follow akun twitter @seputarmasak dan IG kita di https://t.co/YxlS3iT1rU update setiap hari :D	FALSE	0	NA	7/28/2016 15:50	FALSE	NA
2	Hai Ka @adisticorner Suka Masak? follow akun twitter @seputarmasak dan IG kita di https://t.co/YxlS3iT1rU update setiap hari :D	FALSE	0	NA	7/28/2016 15:49	FALSE	NA
3	Hai Ka @adistiayu Suka Masak? follow akun twitter @seputarmasak dan IG kita di https://t.co/YxlS3iT1rU update setiap hari :D	FALSE	0	NA	7/28/2016 15:47	FALSE	NA
4	Hai Ka @adistiarphaa Suka Masak? follow akun twitter @seputarmasak dan IG kita di https://t.co/YxlS3iT1rU update setiap hari :D	FALSE	0	NA	7/28/2016 15:46	FALSE	NA
5	Hai Ka @adistiapratiwii Suka Masak? follow akun twitter @seputarmasak dan IG kita di https://t.co/YxlS3iT1rU update setiap hari :D	FALSE	0	NA	7/28/2016 15:44	FALSE	NA

Fungsi get*User* mempunyai 13 atribut sedangkan *user*Timeline mempunyai 16 atribut, namun tidak semua atribut pada masing-masing fungsi tersebut digunakan dalam penelitian ini. Hasil gabungan atribut antara fungsi get*User* dan fungsi *user*Timeline dapat dilihat pada Tabel 3.3.

Tabel 3. 3 Hasil gabungan atribut antara fungsi get*User* dan fungsi *user*Timeline

Atribut	Fungsi
Nama akun	get*User*
Jumlah *follow*ers	get*User*
Jumlah *follow*ing	get*User*
Jumlah *tweet*	get*User*
Usia akun	get*User*
Rataan *tweet* per hari	get*User*
Verified user	get*User*
rataan selang waktu antar *tweet*	*user*Timeline

3.4 Atribut Data Profil Pengguna (User-based Features)

Pemilihan atribut pada suatu dataset adalah hal yang penting untuk dilakukan pada klasifikasi *spam*. Dengan menganalisa perilaku dari masing-masing akun pengguna *Twitter*, maka akan didapatkan suatu karakteristik antara *spam*mer dan non-*spam*mer.

Ada perbedaan yang amat signifikan antara *spam*mer dan non-*spam*mer. Contohnya hanya sedikit dari *spam*mer, bahkan tidak ada sama sekali, yang menggunakan akun otentik, hal ini dikarenakan mereka tidak menghendaki adanya tanggung jawab sosial pada masa yang akan datang. Sedangkan pengguna pada umumnya akan berupaya untuk mendapatkan akun yang otentik pada *Twitter* untuk menunjukkan reputasi yang mereka miliki kepada publik. Pengguna *Twitter* pada umumnya lebih mengacuhkan pesan *spam* daripada pesan biasanya. *Spam*mer akan mengirim pesan yang sama berulang-ulang kepada lebih banyak pengguna untuk menarik perhatian mereka. Pada penelitian ini akan dijelaskan atribut pengguna berdasarkan perilakunya, meliputi jumlah *tweet*, jumlah *follow*er, jumlah *follow*ing, usia akun, rataan *tweet* per hari, rataan selang waktu antar *tweet*, dan *verified user*.

1. *StatusesCount*/jumlah *tweet* : Atribut ini digunakan untuk mengetahui bahwa seorang pengguna *Twitter* adalah *term*asuk *spam*mer atau bukan. *Spam*mer mengirim duplikat pesan secara massal dalam waktu yang singkat. Mereka mengirim *spam* hanya untuk mendapatkan lebih banyak uang dengan cara menarik perhatian pengguna lain untuk membaca pesan mereka atau mengklik link URL yang mereka tautkan.

2. *Follow*ersCount/jumlah *follow*ers : *Follow*ers adalah pengguna lain yang mengikuti akun Anda di *Twitter*. *Follow*ers akan menerima setiap pesan yang dikirimkan oleh pengguna yang diikuti pada *Twitter*.

3. *FriendsCount*/ jumlah teman : Teman adalah akun yang Anda *follow*. Jika Anda memiliki jumlah "*follow*ing" lebih banyak dari pada jumlah "*follow*ers", maka Anda dapat ditengarai sebagai akun *spam* [5]. Kebiasaan *spam*mer adalah dengan melakukan "*follow*" kepada banyak akun dalam waktu yang singkat. Teman adalah suatu akun yang Anda *follow*/ikuti [3].

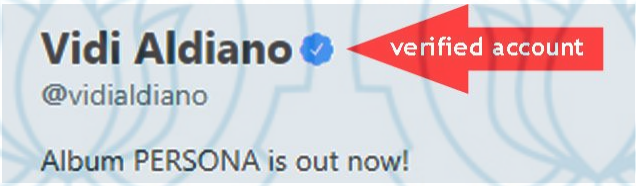
4. *Age of Account*/usia akun : Usia akun adalah tanggal dimana suatu akun pertama kali dibuat.

5. Average *Tweet*s per Day/rataan *tweet* per hari : Atribut ini menghitung rata-rata pesan yang dikirim oleh seorang pengguna dalam sehari. Untuk menghitungnya adalah dengan membagi jumlah *tweet* dengan banyaknya hari sejak akun pertama kali dibuat.

6. Average *limits between Tweet*s/rataan selang waktu antar *tweet*: Penelitian ini mengumpulkan 25 data *tweet*s/pesan terakhir yang dikirim oleh suatu akun. Atribut ini dapat dihitung dengan mengurangkan waktu antara pesan terakhir dan pesan pertama yang dikirim pada *Microblogging Twitter*. Pada umumnya, *spam*mer mempunyai rataan selang waktu antar *tweet* yang lebih kecil dari pada pengguna pada umumnya.

7. *Verified user*/*user* terverifikasi : *User* terverifikasi oleh *Twitter* menunjukkan bahwa akun tersebut adalah akun terpercaya. *Twitter* memberi tanda dengan menambahkan "badge" berbentuk tanda "centang" tepat di sebelah *user*name pada suatu akun *Twitter*. Jika Anda ingin *Twitter* memverifikasi akun Anda, maka Anda dapat mengirim permohonan ke *Twitter* perihal hal tersebut. *Twitter* selanjutnya akan melakukan identifikasi pada akun Anda. Jika syarat dan ketentuan untuk verifikasi akun terpenuhi, maka akun Anda akan diberi tanda centang warna bitu di sebelah *user*name Anda. Tanda centang warna biru ini hanya *Twitter* yang dapat memberikannya, pengguna lain tidak dapat memberi tanda *verified user* dengan sendirinya.

   Akun terverifikasi ini diberikan oleh *Twitter* kepada instansi/lembaga/perorangan yang ahli/berkecimpung dalam berbagai bidang, misalnya bidang pendidikan, hiburan/entertainment, agama, music, pemerintahan, tokoh masyarakat, artis/aktris, lembaga pemerintah, situs berita, dll. Akun terverifikasi oleh *Twitter* haruslah akun yangdiakui oleh publik. Salah satunya adalah jumlah *follow*ers yang melebihi jumlah *follow*ers yang dipunyai oleh akun pada umumnya. Akun yang diberi "*verified* badge" bukan berarti akun ini di-endorse oleh *Twitter* [20]. Contoh akun yang telah diverifikasi oleh *Twitter* dapat dilihat pada Gambar 3.3.

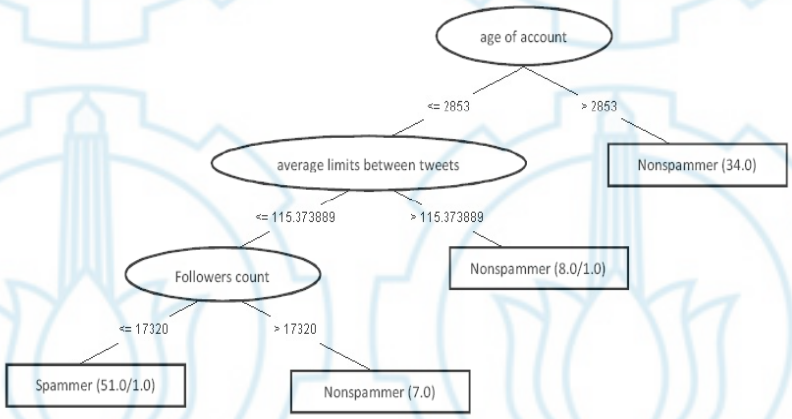


Gambar 3. 4 Contoh akun yang telah diverifikasi oleh *Twitter*

3.5 Model Pohon Keputusan (Decision Tree)

Data yang telah melalui tahap praproses dijadikan bahan masukan pada sistem klasifikasi yang menggunakan metode Decision Tree. Hasil yang didapatkan adalah model klasifikasi berupa pohon keputusan (tree). Jumlah data yang diteliti adalah 100 akun *Twitter*. Akun *spam*mer berjumlah 51 akun, sedangkan akun non-*spam*mer berjumlah 49 akun. Akun non-*spam*mer terdiri dari 25 akun pengguna

biasa dan 24 akun lembaga/instansi/akun ter-verifikasi *Twitter*. Model pohon keputusan (tree) untuk klasifikasi *spam*mer dan non-*spam*mer dapat dilihat pada Gambar 3.4.



Gambar 3. 5 Model pohon keputusan (tree) untuk klasifikasi *spam*mer dan non-*spam*mer

Klasifikasi dengan menggunakan metode Decision Tree ini menghasilkan model pohon keputusan (tree) dengan aturan :

1. Jika usia akun >2853 hari, maka diklasifikasikan sebagai kelas non-*spam*mer
2. Jika rataan selang waktu antar *tweet*s >115,373889, maka diklasifikasikan sebagai kelas non-*spam*mer
3. Jika jumlah *follow*ers >17320, maka diklasifikasikan sebagai kelas non-*spam*mer

Pada Gambar 3.4 menunjukkan bahwa pada model klasifikasi berupa pohon keputusan (tree) ini, atribut usia akun mempunyai peluang yang paling besar dalam pengambilan keputusan. Atribut usia akun menandakan bahwa setiap pengecekan data harus melalui atribut usia akun terlebih dahulu sebelum kemudian dilakukan pengecekan pada atribut lainnya. Jadi atribut usia akun memiliki peran penting daripada atribut yang lainnya.

Saat pengecekan data, sistem akan melakukan cek terhadap usia akun terlebih dahulu, jika usia akun ≤2853 hari, langkah selanjutnya adalah dilakukan cek atribut rataan selang waktu antar *tweet*s, setelah itu dilakukan cek atribut jumlah *follow*ers. Tahap terakhir adalah melakukan pengecekan terhadap data dengan atribut usia akun >2853 hari.

Klasifikasi *spam*mer tidak memerlukan tujuh atribut yaitu jumlah *tweet*, jumlah *follow*er, jumlah *follow*ing, usia akun, rataan *tweet* per hari, rataan selang waktu antar *tweet*, dan *verified user*, namun, pada model pohon keputusan ini dapat dianalisa bahwa cukup dengan menggunakan tiga atribut saja yaitu usia akun, rataan selang waktu antar *tweet*s, dan jumlah *follow*ers maka sistem dapat mengklasifikasikan antara *spam*mer dan non-*spam*mer.

Hal ini dikarenakan akun non-*spam*mer mempunyai usia yang lebih panjang dari pada akun *spam*mer, selain itu, akun non *spam*mer mempunyai angka yang lebih tinggi dalam atribut rataan selang waktu antar *tweet* yang artinya angka ini harus tinggi karena non-*spam*mer tidak terlalu focus untuk menyebarkan pesan secara massal dalam waktu yang singkat seperti yang *spam*mer lakukan, sehingga angka *spam*mer dalam hal rataan selang waktu antar *tweet* lebih rendah dari non-*spam*mer. Di samping itu, jumlah *follow*ers akun non-*spam*mer adalah lebih tinggi dari pada akun *spam*mer, yaitu >17320 *follow*ers.

3.6 Desain Sistem Naïve Bayes

Pada tahap ini, metode yang digunakan dalam pengklasifikasian *spam*mer berdasarkan konten *tweet*s pengguna (*content-based features*) adalah Naïve Bayes. Secara umum proses ini dibagi menjadi beberapa tahap. Tahap-tahap tersebut dapat dilihat pada Gambar 3.6.

Dalam klasifikasi dokumen berbasis teks, dokumen dapat diklasifikasikan dengan mengevaluasi kata yang terdapat di dalamnya. Naive Bayes adalah salah satu metode yang popular yang dapat digunakan untuk mengklasifikasikan dokumen secara probabilistic. Metode Naive Bayes adalah jenis dari Teori Bayesian dimana kondisi atau kelas yang ada adalah independen dan tidak terikat satu sama lain. Kata yang membentuk sebuah dokumen di kategori tertentu tidak

mempengaruhi kategori lain. Oleh karena itu, probabilitas dari dokumen yang diuji akan dihitung masing-masing.

Tahap-tahap pengklasifikasian *spam*mer menggunakan Naïve Bayes adalah sebagai berikut :

1. Probabilitas dokumen dari kata yg menyusun (W1…W2) memiliki kelas tertentu (C) dapat dihitung menggunakan persamaan 3. 5

$$P(C|W1,\dots,Wn) = P(C)\prod_{i=1}^{n} P(W1|C) \quad (3.5)$$

2. Kemungkinan untuk kata berada kategori tertentu diperoleh dari pembagian antara frekuensi kemunculan kata dalam sebuah dokumen (nk), jumlah kata dalam kategori yang diamati (nc), dan jumlah variasi kata (n) dari semua dokumen dalam data pelatihan untuk kategori tertentu seperti pada persamaan 3. 6

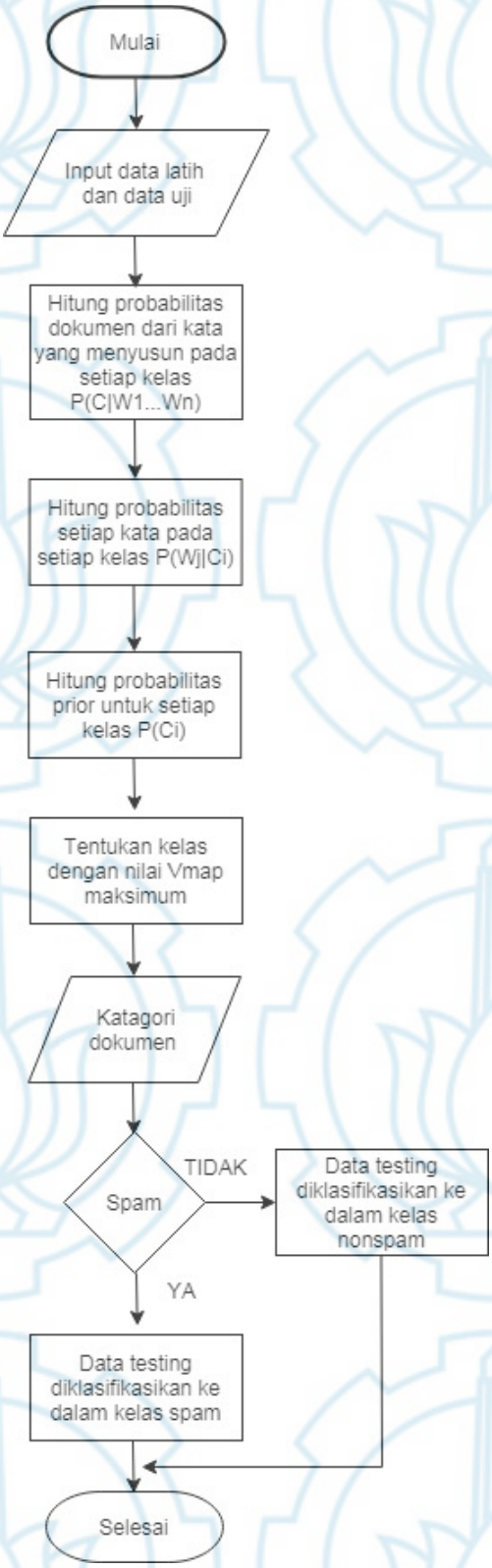
$$P(Wj|Ci) = (1 + n_k)/\ (n_c + n) \quad (3.6)$$

3. Probabilitas prior untuk kategori tertentu diperoleh dari jumlah dokumen (nd) dalam kategori tertentu dibagi dengan jumlah total dokumen (nt) dari semua kategori dalam data pelatihan yang diberikan seperti pada persamaan 3. 7

$$P(Ci) = |n_d|/\ |n_t| \quad (3.7)$$

4. Setelah nilai-nilai posterior dari dokumen dievaluasi terhadap setiap kategori telah dihitung, dokumen tersebut diklasifikasikan berdasarkan nilai maksimum-a-posteriori (VMAP) seperti pada persamaan 3. 8. Oleh karena itu, dokumen diklasifikasikan ke dalam kategori yang mana nilai posteriornya paling besar.

$$V_{map = argmax}\ P(Ci)\prod P(Wj|Ci) \quad (3.8)$$



Gambar 3. 6 Desain sistem Naïve Bayes

3.7 Praproses Data *Spam*mer Berdasarkan Konten *Tweet* (Content-based Features) menggunakan Metode Naïve Bayes

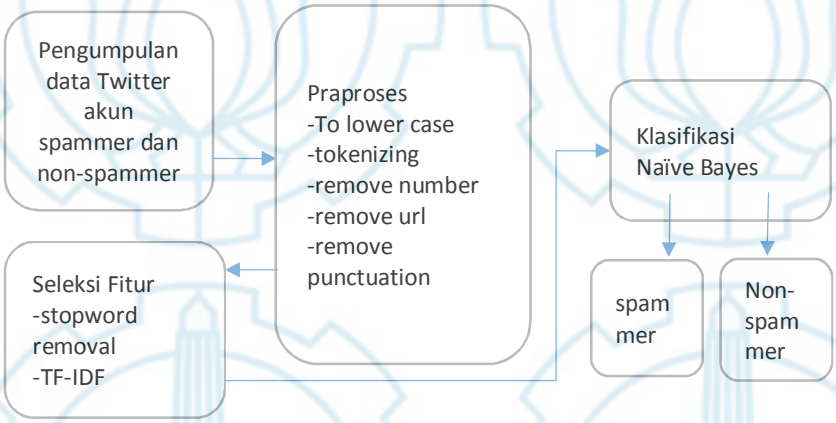
Data mining adalah disiplin ilmu yang tujuan utamanya adalah untuk menambang pengetahuan dari data atau informasi yang dimiliki. Text mining adalah salah satu solusi yang dapat membantu permasalahan diatas [21]. Text mining mirip dengan data mining, kecuali pada teknik data mining yang didesain untuk pengerjaan data yang terstruktur pada sebuah database, tapi text mining dapat bekerja pada data yang tidak terstruktur atau semi testruktur seperti email, sebuah dokumen text lengkap, html dan lain-lain. Sehingga text mining merupakan sebuah penemuan baru dari informasi yang belum diketahui dengan mengekstrak informasi dari sumber tertulis.

Blok diagram sistem klasifikasi *spam*mer dan non-*spam*mer dapat dilihat pada Gambar 3.5. Menurut [21] ada beberapa langkah yang dilakukan dalam text mining :

1. Text Preprocessing

   Tindakan yang dilakukan pada tahap ini adalah:

   1. *To lower case*, yaitu mengubah semua karakter huruf menjadi huruf kecil.
   2. *Tokenizing*, yaitu proses penguraian deskripsi yang semula berupa kalimat – kalimat menjadi kata-kata.
   3. *Remove number*, yaitu menghilangkan karakter angka yang ada pada kata tersebut.
   4. *Remove url*, yaitu menghilangkan link internet.
   5. *Remove punctuation*, yaitu menghilangkan delimiter-delimiter seperti tanda titik(.), koma(,) dan spasi.



Gambar 3. 7 Blok diagram sistem klasifikasi *spam*mer dan non-*spam*mer berdasarkan konten *Tweet*

2. Feature Selection

Pada tahap ini tindakan yang dilakukan adalah:

1. *Stopword (stopword removal*) adalah kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen. Stopword untuk bahasa Indonesia diperoleh dari: http://www.ranks.nl/stopwords/indonesian. Contoh stopword dapat dilihat pada Tabel 4.1

Tabel 4. 1 Contoh data stopword

Id_stopword	katastopword
1	Ada
2	Adalah
3	adanya

2. Pembobotan kata (TF-IDF)

Dalam analisis dokumen *tweet*s, pembobotan kata digunakan untuk mendapatkan suatu topik atau keyword dari kumpulan *tweet*s. Salah satu metode pembobotan adalah TF-IDF (*Term* Frequency – Inverse Document Frequency).

Nilai bobot suatu kata (*term*) menyatakan kepentingan bobot tersebut dalam merepresentasikan *tweet*s. Pada pembobotan TF-IDF, bobot akan semakin besar jika frekuensi kemunculan kata semakin tinggi, tetapi bobot akan berkurang jika kata tersebut semakin sering muncul pada *tweet*s lainnya

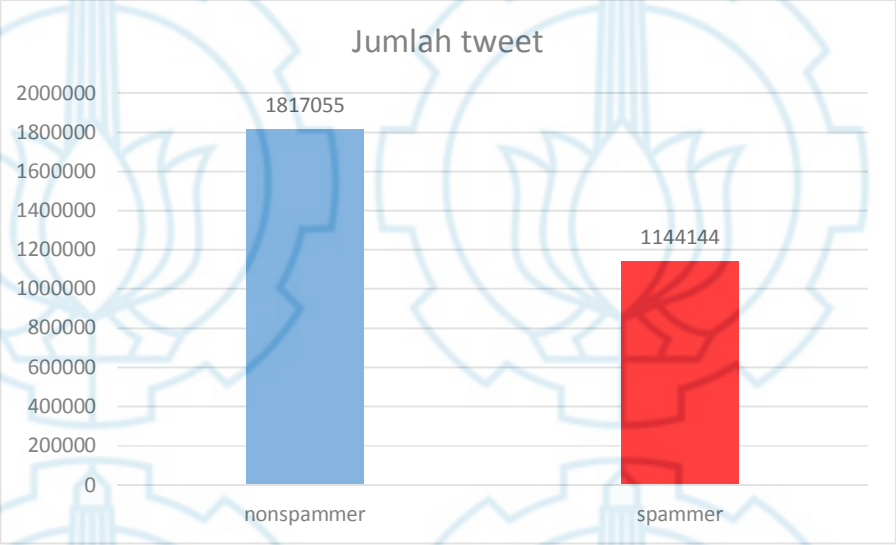
Tahapan selanjutnya adalah membentuk *Term* Document Matrix (TDM). TDM menunjukkan hubungan antara *term* dan dokumen, dimana setiap baris berisi *term* dan setiap kolom untuk dokumen. TDM merepresentasikan jumlah kemunculan suatu kata pada dokumen

Halaman ini sengaja dikosongkan

BAB 4
HASIL DAN PEMBAHASAN

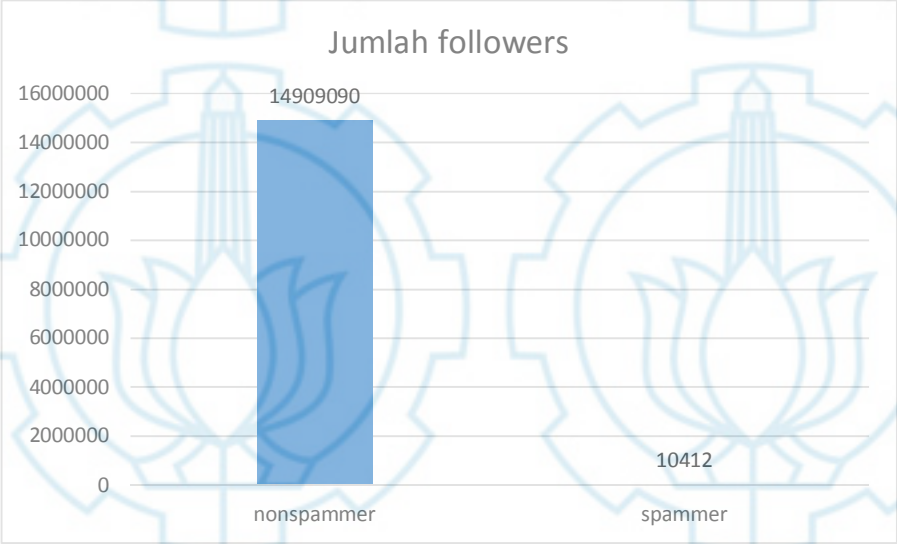
4.1 Analisa *Spam*mer Berdasarkan Profil Pengguna (*User*-based Features) menggunakan Metode Decision Tree

Data yang digunakan pada penelitian ini adalah sebanyak 100 data yang merupakan akun pengguna *Twitter* yang dikumpulkan melalui API *Twitter*. Data tersebut terdiri dari akun *spam*mer berjumlah 51 akun, sedangkan akun non-*spam*mer berjumlah 49 akun. Akun non-*spam*mer terdiri dari 25 akun pengguna biasa dan 24 akun lembaga/instansi/akun ter-verifikasi *Twitter*. Masing-masing akun diberi atribut berdasarkan profil akun, meliputi jumlah *tweet*, jumlah *follow*er, jumlah *follow*ing, usia akun, rataan *tweet* per hari, rataan selang waktu antar *tweet*, dan *verified user*. Tahap selanjutnya adalah memberi label kelas *spam*mer dan non-*spam*mer. Penelitian ini menggunakan metode Decision tree untuk melakukan klasifikasi apakah suatu akun merupakan *spam*mer atau tidak.



Gambar 4. 1 Jumlah *tweet*

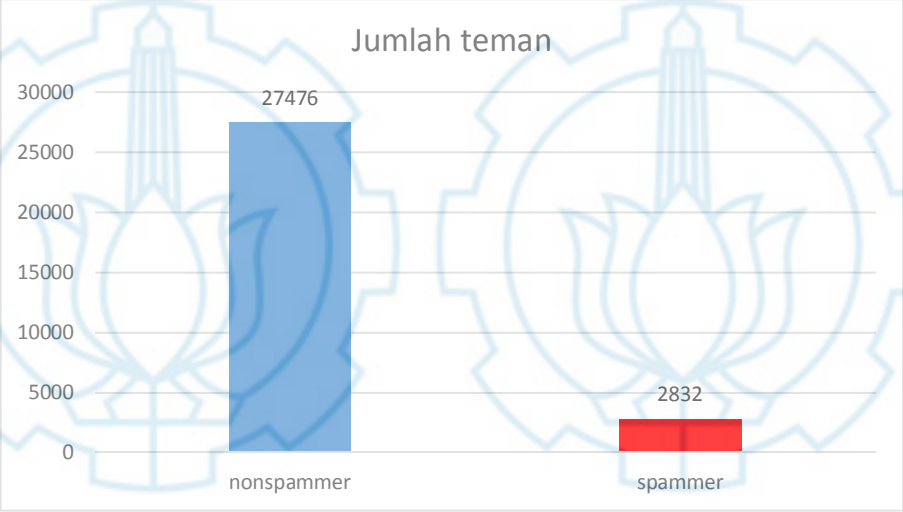
Gambar 4.1 menunjukkan jumlah *tweet*s antara *spam*mer dan non-*spam*mer. Dari teori sebelumnya, bahwa *spam*mer mempunyai banyak *tweet*s (pesan) daripada non-*spam*mer, maka pada grafik ini menunjukkan hal yang berbeda. Non-*spam*mer mempunyai jumlah *tweet*s lebih besar dari pada jumlah *tweet*s pada *spam*mer. Jumlah *tweet*s pada non *spam*mer mencapai 1817055, sedangkan jumlah *tweet*s pada *spam*mer mencapai 1144144. Setelah dicek kembali ternyata akun non-*spam*mer yang mempunyai status/jumlah *tweet* tertinggi adalah dari akun salah satu situs layanan pelanggan @indosatcare. Situs ini melakukan layanan kepada pelanggan secara massal dan harus menjawab pertanyaan pelanggan yang banyak menggunakan produk dari mereka. Sehingga jumlah *tweet*snya lebih tinggi daripada *spam*mer. Akun ini juga merupakan akun terverifikasi oleh *Twitter*. Jadi tidak heran jika masyarakat mempercayai untuk menjadi *follow*ers pada akun @indosatcare.



Gambar 4. 2 Jumlah *follow*ers

Jumlah *follow*ers pada Grafik 4.2 menunjukkan bahwa untuk akun non-*spam*mer mempunyai jumlah *follow*ers tertinggi, yaitu mencapai 14909090. Sedangkan untuk akun *spam*mer menunjukkan angka 10412 *follow*ers. Masyarakat

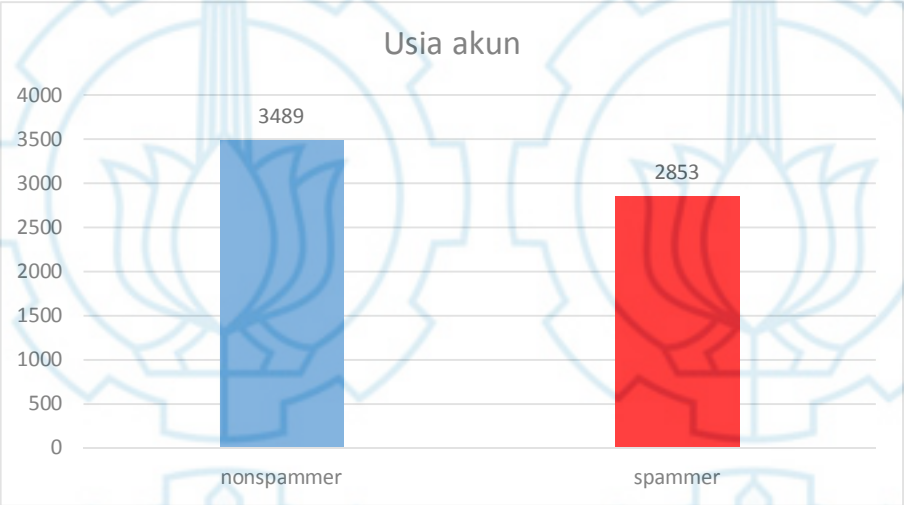
menjadi lebih cerdas seiring perkembangan waktu, untuk dapat mengambil keputusan akun mana yang layak diikuti dan akun mana yang sifatnya mengganggu sehingga patut diabaikan. Tidak semua *spam*mer mempunyai jumlah *follow*ers yang tinggi, sehingga dalam hal ini tidak dapat dijadikan patokan jika jumlah *follow*ernya tinggi maka akun tersebut adalah akun *spam*mer.



Gambar 4. 3 Jumlah teman

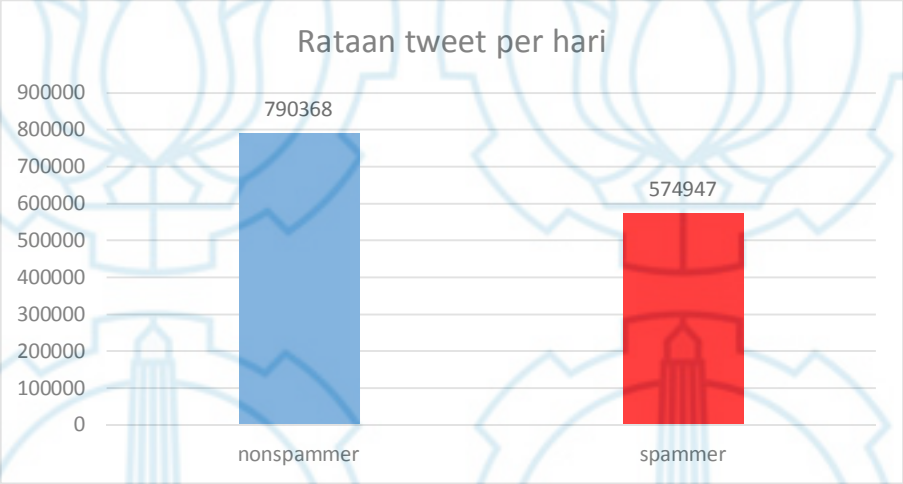
Gambar 4.3 menunjukkan jumlah teman/jumlah *follow*ing antara akun *spam*mer dan non-*spam*mer. Jumlah teman dari akun non-*spam*mer sangat signifikan daripada akun *spam*mer. Jumlah teman pada akun *spam*mer jauh lebih kecil dari dapa akun non-*spam*mer. Hal ini ditunjukkan bahwa akun non-*spam*mer mempunyai jumlah teman 274760, sedangkan akun *spam*mer hanya mencapai 2832. Seperti yang telah dijelaskan pada bab sebelumnya, teman adalah akun yang Anda *follow*. Jika Anda memiliki jumlah "*follow*ing" lebih banyak dari pada jumlah "*follow*ers", maka Anda dapat ditengarai sebagai akun *spam* [5]. Kebiasaan *spam*mer adalah dengan melakukan "*follow*" kepada banyak akun dalam waktu yang singkat. Teman adalah suatu akun yang Anda *follow*/ikuti [3]. Dalam hal ini *spam*mer tidak perlu mengikuti akun pengguna lain untuk menyebarkan pesan *spam*nya (*spam tweet*s). *Spam*mer hanya perlu menyebutkan nama (*user*name) pengguna lain dengan menambahkan "@namapenggunalain" pada setiap posting

untuk menarik perhatian pengguna tersebut. Sehingga pengguna yang disebutkan itu, mau tidak mau, langsung atau secara tidak langsung, membaca pesan yang disebarkan *spam*mer. Hasil dari penelitian ini menunjukkan bahwa *spam*mer mempunyai jumlah teman yang kecil dan *follow*ers relatif banyak. Sementara tidak menutup kemungkinan beberapa *spam*mer mempunyai jumlah *follow*ers dan jumlah teman yang relatif banyak.



Gambar 4. 4 Usia Akun

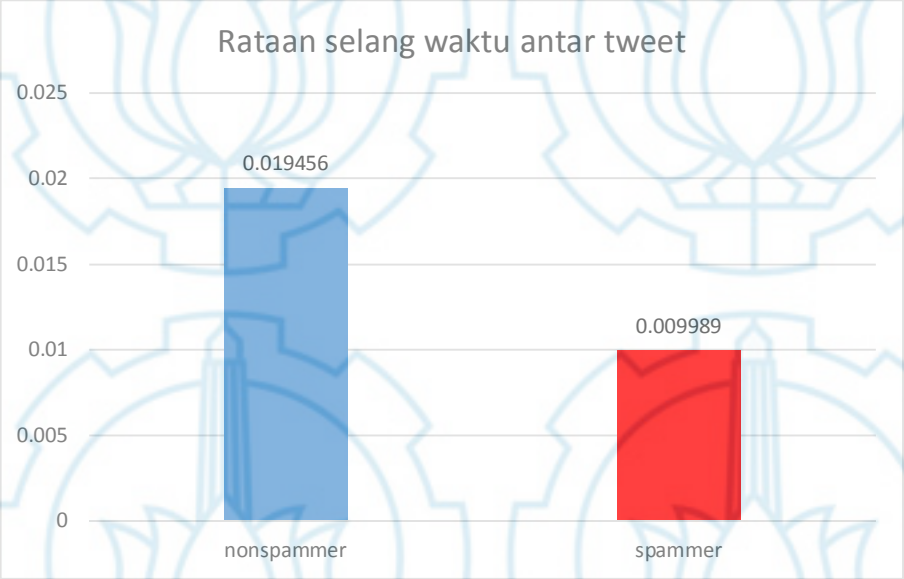
Gambar 4.4 menunjukkan bahwa usia akun dari non-*spam*mer dan *spam*mer tidak terlalu jauh. Untuk usia akun non-*spam*mer adalah 3489 hari, sedangkan usia akun *spam*mer 2853 hari. Usia akun non-*spam*mer lebih lama dibandingkan usia akun *spam*mer. Hal ini disebabkan *spam*mer akan membuat akun duplikat karena mereka selalu dilaporkan oleh pengguna lain sebagai *spam* kepada *Twitter*, selain itu mereka juga sering di-block oleh pengguna lainnya karena pesan yang mengganggu, sehingga akun lama dibiarkan begitu saja. Beberapa *spam*mer mempunyai banyak duplikat akun, pada data yang telah terkumpul, terdapat *spam*mer dengan *user*name berbeda, namun konten dari *tweet*s mereka sama. Hasil penelitian ini menunjukkan bahwa semua *spam*mer mempunya usia akun yang pendek sementara pada dataset kami ada beberapa *spam*mer yang hanya berusia beberapa bulan saja.



Gambar 4. 5 Rataan *Tweet* per Hari

Gambar 4.5 menunjukkan rataan *tweet* per hari yang paling tinggi adalah pada akun non-*spam*mer. Akun non-*spam*mer mempunyai rataan *tweet* per hari mencapai 790369. Sedangkan pada akun *spam*mer mencapai 574947. Kami analisa bahwa akun non-*spam*mer yang mempunyai rataan *tweet* per hari tertinggi adalah akun sebuah situs layanan pelanggan @indosatcare. Informasi pada situs ini amat berguna bagi masyarakat pengguna produk mereka. Keluha dan pertanyaan pelanggan selalu diutamakan untuk dijawab pada setiap pesan yang dikirimkan untuk dibaca pelanggan yang bersangkutan maupun pelanggan lainnya yang mempunyai kasus/pertanyaan yang sama. Sedangkan akun *spam*mer tidak mempunyai update pesan terbaru, dengan kata lain, *spam*mer hanya berkutat pada pesan yang sama di waktu yang berbeda sekalipun. Itulah sebabnya mengapa rataan *tweet* per hari akun *spam*mer tidak terlalu tinggi.

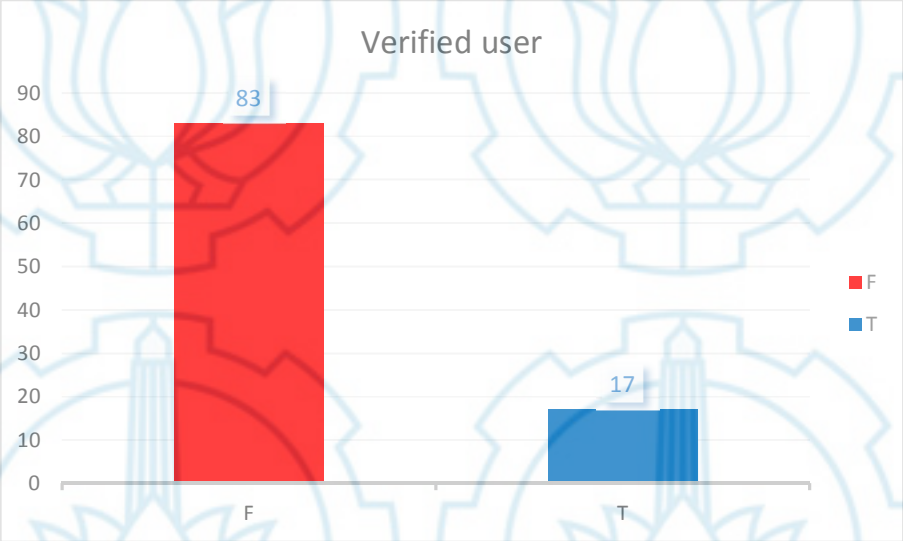
Pembahasan selanjutnya adalah tentang rataan selang waktu antar *tweet*. Pada Gambar 4.6 menunjukkan bahwa rataan selang waktu antar *tweet* untuk *spam*mer adalah yang palin kecil dibandingkan dengan non-*spam*mer. Penelitian ini mengumpulkan 25 data *tweet*s/pesan terakhir yang dikirim oleh suatu akun. Rataan selang waktu antar *tweet* dapat dihitung dengan mengurangkan waktu antara pesan terakhir dan pesan pertama yang dikirim pada *Microblogging Twitter*. Pada umumnya, *spam*mer mempunyai rataan selang waktu antar *tweet* yang lebih kecil dari pada pengguna pada umumnya.



Gambar 4. 6 rataan selang waktu antar *tweet*

Pada kenyataannya memang demikian, *spam*mer mempunyai nilai rataan selang waktu antar *tweet* sebesar 0,009989 sedangnkan non-*spam*mer mempunyai nilai rataan selang waktu antar *tweet* sebesar 0, 019456. Beberapa *spam*mer tidak mem-*follow* banyak pangguna namun hanya focus menyebarkan pesan *spam* secara massal setelah akun selesai dibuat saat itu juga [2]. Untuk akun non-*spam*mer yang mempunyai nilai rataan selang waktu antar *tweet* terkecil kedua setelah akun *spam*mer adalah akun situs berita @kompascom. Situs ini selalu update isi beritanya bahkan setiap detik. Jadi tidak heran jika nilai rataan selang waktu antar *tweet*nya terkecil kedua setelah akun *spam*mer.

User terverifikasi oleh *Twitter* menunjukkan bahwa akun tersebut adalah akun terpercaya. *Twitter* memberi tanda dengan menambahkan "badge" berbentuk tanda "centang" tepat di sebelah *user*name pada suatu akun *Twitter*. Jika Anda ingin *Twitter* memverifikasi akun Anda, maka Anda dapat mengirim permohonan ke *Twitter* perihal hal tersebut. *Twitter* selanjutnya akan melakukan identifikasi pada akun Anda. Jika syarat dan ketentuan untuk verifikasi akun terpenuhi, maka akun Anda akan diberi tanda centang warna bitu di sebelah *user*name Anda. Tanda centang warna biru ini hanya *Twitter* yang dapat memberikannya, pengguna lain tidak dapat memberi tanda *verified user* dengan sendirinya.



Gambar 4. 7 *Verified user*

Gambar 4.7 menunjukkan bahwa semua akun *spam*mer bukan merupakan akun terpercaya, bukan akun yang terverifikasi oleh *Twitter*. Berbeda dengan akun non-*spam*mer, meskipun tidak semua akun non-*spam*mer diverifikasi oleh *Twitter*, namun ada 17 akun non-*spam*mer yang mempunyai tanda centang biru. Akun terverifikasi ini diberikan oleh *Twitter* kepada instansi/lembaga/perorangan yang ahli/berkecimpung dalam berbagai bidang, misalnya bidang pendidikan, hiburan/entertainment, agama, music, pemerintahan, tokoh masyarakat, artis/aktris, lembaga pemerintah, situs berita, dll. Akun terverifikasi oleh *Twitter* haruslah akun yangdiakui oleh publik. Salah satunya adalah jumlah *follow*ers yang melebihi jumlah *follow*ers yang dipunyai oleh akun pada umumnya. Akun yang diberi "*verified* badge" bukan berarti akun ini di-endorse oleh *Twitter* [19].

Hasil analisa klasifikasi *spam*mer dan non-*spam*mer berdasarkan profil pengguna, yang meliputi jumlah *tweet*, jumlah *follow*er, jumlah *follow*ing, usia akun, rataan *tweet* per hari, rataan selang waktu antar *tweet*, dan *verified user*. dapat dilihat pada Tabel 4.2.

Tabel 4. 2 Hasil analisa berdasarkan profil pengguna

Field Attributes	Account Class	
	*Spam*mer	Non-*spam*mer
StatusesCount	1.144.144	1.817.055
*Follow*ersCount	10.412	14.909.090
FriendsCount	2.832	274.760
Age of account	2.853	3.489
Average *tweet*s per day	574.947	790.368
Average limits between *tweet*s	0.009989	0.019456
Verified user	False	True, False

4.2 Hasil Uji Coba Berdasarkan Profil Pengguna (*User*-based Features) menggunakan Metode Decision Tree

Pada pengujian klasifikasi *spam*mer dan non-*spam*mer berdasarkan profil pengguna didapatkan hasil yang ditunjukkan pada Tabel 4.3.

Tabel 4. 3 Hasil uji coba berdasarkan profil pengguna

		Prediction	
		*Spam*mer	Non-*spam*mer
True	*Spam*mer	12	3
	Non-*spam*mer	1	18

Tabel 4.3 menunjukkan bahwa 12 akun *spam*mer diklasifikasikan secara benar sebagai *spam*mer. Dan 18 akun non-*spam*mer juga diklasifikasikan secara benar sebagai kelas non-*spam*mer. Hanya 1 akun non-*spam*mer yang diklasifikasikan sebagai akun *spam*mer. Dan ada 3 akun *spam*mer yang diklasifikasikan bukan pada kelasnya.

Perhitungan Accuration, Precision, Sensitivity, dan Specitivity adalah sebagai berikut :

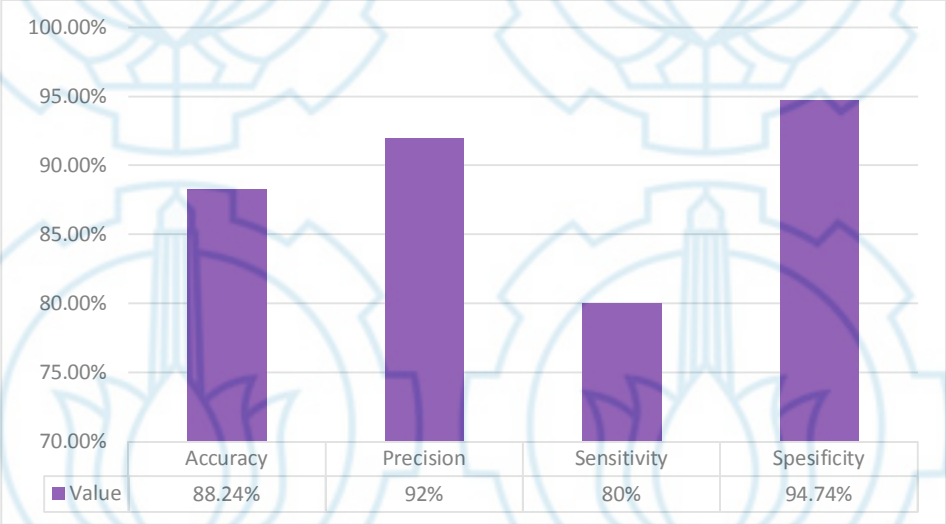
$$\text{Accuracy} = \frac{12+18}{12+18+1+3} \text{ x100\%} = 88.235\%$$

$$\text{Precision} = \frac{12}{12+1} \times 100\% = 92\%$$

$$\text{Sensitivity} = \frac{12}{12+3} \times 100\% = 80\%$$

$$\text{Specificity} = \frac{18}{18+1} \times 100\% = 94,74\%$$

Dari perhitungan tersebut di atas dapat dibuat grafik sebagai berikut :



Gambar 4. 8 Grafik kinerja sistem

Gambar 4.8 menunjukkan hasil dari klasifikasi *spam*mer dan non-*spam*mer menggunakan metode Decision Tree dengan split 66.0% train. Hasil accuracy sebesar 88,24%, precision 92%, sensitivity 80% dan specificity sebesar 94,74%. Hal ini menunjukkan bahwa sistem dapat mengenali akun non-*spam*mer dengan lebih baik daripada akun *spam*mer. Hal ini didukung oleh prosentase specificity mempunyai nilai yang paling besar.

4.3 Analisa *Spam*mer Berdasarkan Konten *Tweet*s (Content-based Features) menggunakan Metode Naïve Bayes

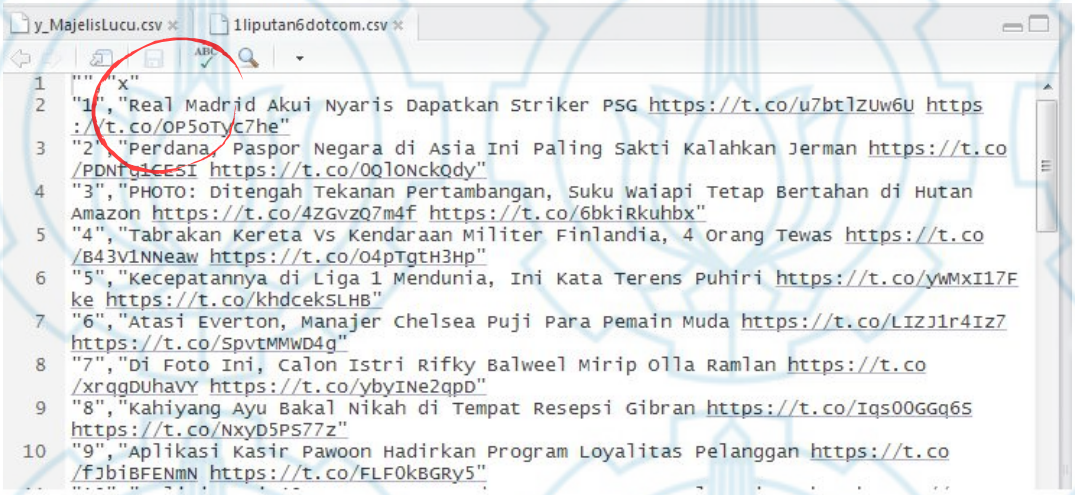
Data yang digunakan pada penelitian ini adalah sebanyak 100 data yang merupakan akun pengguna *Twitter* yang dikumpulkan melalui API *Twitter*. Data tersebut terdiri dari akun *spam*mer berjumlah 51 akun, sedangkan akun non-*spam*mer berjumlah 49 akun. Akun non-*spam*mer terdiri dari 25 akun pengguna biasa dan 24 akun lembaga/instansi/akun ter-verifikasi *Twitter*. Kami juga mengumpulkan 2156 *tweet* terakhir dari masing-masing akun *spam*mer dan non-*spam*mer.

Analisis data menggunakan metode text mining dengan bantuan software *Rstudio*. Package yang digunakan adalah t*witteR*, ggplot2, wordcloud, tm. Adapun metode analisis yang digunakan untuk mencapai tujuan penelitian dalam penelitian ini diuraikan sebagai berikut:

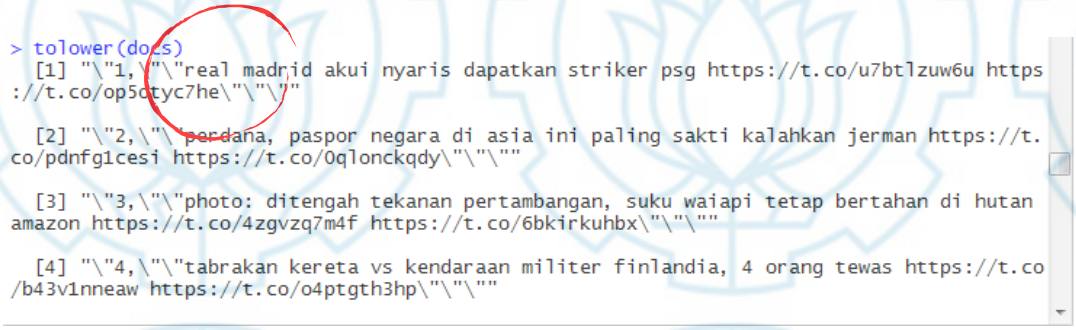
1. Membuat akun pada API *Twitter*, untuk memperoleh *consumer key*, *consumer secret, access token, dan access token secret* yang akan digunakan untuk mengambil data text *Twitter* dengan software *Rstudio*.
2. Text praproses, dimana data teks yang telah diambil dari *Twitter* diolah melalui beberapa tahap, yaitu:
    a. *To lower case*
    b. *Tokenizing*
    c. *Remove number*
    d. *Remove url*
    e. *Remove punctuation*
3. Feature selection, dimana data text yang telah melalui tahap text praproses dilakukan proses selanjutnya, yaitu:
    a. *Stopword (stopword removal)*
    b. Pembobotan kata (TF-IDF)
4. Data text yang telah disusun ulang, kemudian dibuat *term*-document matrix. Matrix yang terbentuk merupakan matrik yang telah diberi pembobotan TF-IDF.
5. Membuat barplot dan wordcloud dari *term*-document matrix dengan pembobotan TF-IDF.

6. Menentukan jumlah *term* frekuensi

Setelah terkumpul data teks sebanyak 2156 *tweet*, maka tahap selanjutnya adalah melakukan praproses teks. Contoh data *tweet* sebelum dilakukan praproses teks dapat dilihat pada Gambar 4.9. Sedangkan contoh praproses teks pada segmen to lower case dapat dilihat pada gambar 4.10, dan contoh praproses teks pada segmen remove punctuation dapat dilihat pada gambar 4.11.



Gambar 4. 9 Contoh data *tweet* sebelum praproses teks



Gambar 4. 10 Contoh praproses teks to lower case

```
> docs2 <- gsub("[[:digit:]]", " ", docs2)
> docs3 <- gsub("[[:digit:]]", " ", docs2)
> stripWhitespace(docs2)
  [1] " real madrid akui nyaris dapatkan striker psg https t co u btlzuw u https t co op
otyc he "
  [2] " perdana paspor negara di asia ini paling sakti kalahkan jerman https t co pdnfg c
esi https t co qlonckqdy "
  [3] " photo ditengah tekanan pertambangan suku waiapi tetap bertahan di hutan amazon ht
tps t co zgvzq m f https t co bkirkuhbx "
  [4] " tabrakan kereta vs kendaraan militer finlandia orang tewas https t co b v nneaw h
ttps t co o ptgth hp "
  [5] " kecepatannya di liga mendunia ini kata terens puhiri https t co ywmxi fke https t
 co khdceks lhb "
  [6] " atasi everton manajer chelsea puji para pemain muda https t co lizj r iz https t
co spvtmmwd g "
  [7] " di foto ini calon istri rifky balweel mirip olla ramlan https t co xrqqduhavy htt
```

Gambar 4. 11 Contoh praproses teks remove punctuation

Pada Gambar 4.10 dan 4.11 dapat dilihat bahwa huruf besar kecil harus diubah menjadi huruf kecil semua, sedangkan semua tanda baca, angka, URL, *hashtag* dan *mention* juga harus dihilangkan.

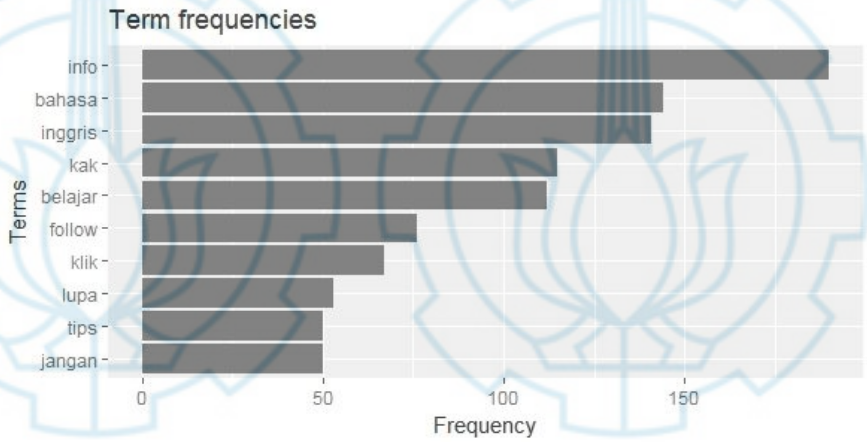
adisti	adistia	adistiaaw	adistiabia	adistiades	adistiadis	adistiaghi	adistianid	adistianti	adistiapra	adistiarph	adistiayu	adisticorn	adistidisti
0	0	0	0	0	0	0	0	0	0	0	0	0	0.928083
0	0	0	0	0	0	0	0	0	0	0	0	0.928083	0
0	0	0	0	0	0	0	0	0	0	0	0.928083	0	0
0	0	0	0	0	0	0	0	0	0	0.928083	0	0	0
0	0	0	0	0	0	0	0	0	0.928083	0	0	0	0
0	0	0	0	0	0	0	0	0.928083	0	0	0	0	0
0	0	0	0	0	0	0	0.928083	0	0	0	0	0	0
0	0	0	0	0	0	0.928083	0	0	0	0	0	0	0
0	0	0	0	0	0.928083	0	0	0	0	0	0	0	0
0	0	0	0	0.928083	0	0	0	0	0	0	0	0	0
0	0	0	0.928083	0	0	0	0	0	0	0	0	0	0
0	0	0.928083	0	0	0	0	0	0	0	0	0	0	0
0	0.928083	0	0	0	0	0	0	0	0	0	0	0	0
0.734589	0	0	0	0	0	0	0	0	0	0	0	0	0
0.678082	0	0	0	0	0	0	0	0	0	0	0	0	0
0.678082	0	0	0	0	0	0	0	0	0	0	0	0	0
0.678082	0	0	0	0	0	0	0	0	0	0	0	0	0
0.734589	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0

Gambar 4. 12 . Contoh *term* document matrix dengan pembobotan TF-IDF

Setelah semua tahap praproses teks dilakukan, langkah selanjutnya adalah menyusun ulang data teks dan membuat *term* dokumen matrix dengan pembobotan TF-IDF. Dalam hal ini terdapat 2156 dokumen dengan 1471 atribut. Atribut sebanyak 1471 tersebut diperoleh dari kata-kata (*term*) yang dihasilkan setelah tahap praproses. *Term* tersebut menjadi atribut pada proses klasifikasi *spam* dan non*spam* dengan metode Naïve Bayes. *Term* ini yang nantinya dihitung nilai

maksimum-a-posteriori (VMAP) untuk menentukan apakah dokumen ini *term*asuk *spam* atau bukan. Contoh *term* document matrix dengan pembobotan TF-IDF dapat dilihat pada Gambar 4.12.

Dalam tahap praproses dihasilkan kata –kata yang disebut token. Token selanjutnya akan menjadi *term*. *Term* adalah token unik. Frekuensi kemunculan suatu *term* dapat dilihat pada Gambar 4.13.



Gambar 4. 13 Frekuensi kemunculan suatu *term*

Pada Gambar 4.13 dapat dilihat bahwa *term* yang paling sering muncul adalah "info" dengan kemunculan lebih dari 150 kali. Artinya pada sekumpulan data *tweet*s cenderung berisi informasi *spam* yang cenderung diulang-ulang secara massal mengenai info belajar Bahasa Inggris yang mudah dan murah. Rata-rata *spam*mer mengirim duplikat pesan yang sama secara massal kepada semua pengguna *Twitter*. Selain mengenai info belajar Bahasa Inggris, kecenderungan lainnya adalah selalu saja *spam*mer menyarankan pengguna lainnya untuk *follow* suatu akun tertentu atau meng klik laman tertentu. Sehingga pengguna lain tertarik untuk mem *follow* atau meng klik link URL yang telah ditautkan. Info yang lain adalah tips kesehatan. Sehingga *term* "tips" juga *term*asuk dalam 10 *term* teratas untuk kemunculan yang paling sering.

4.4 Hasil Uji Coba Berdasarkan Konten *Tweet*s (Content-based Features) menggunakan Metode Naïve Bayes

Setelah tahap *term* frekuensi selesai, maka dilakukan evaluasi data. Dari 2156 *tweet* dibagi menjadi data train sebanyak 1558 dan data test sebanyak 598. Hasil dari klasifikasi *spam*mer dan non-*spam*mer berdasarkan konten *tweet* menggunakan metode Naïve Bayes pada data test dapat dilihat pada tabel 4.4

Tabel 4. 4 Hasil uji coba berdasarkan konten *tweet*

		Prediction	
		*Spam*mer	Non-*spam*mer
True	*Spam*mer	388	13
	Non-*spam*mer	15	182

Dari tabel dapat dilihat bahwa sebanyak 570 data dapat diklasifikasikan secara benar menurut kelasnya. Dan hanya 28 data yang diklasifikasikan tidak pada tempatnya. Tabel 4.4 menunjukkan bahwa 388 akun *spam*mer diklasifikasikan secara benar sebagai *spam*mer. Dan 182 akun non-*spam*mer juga diklasifikasikan secara benar sebagai kelas non-*spam*mer. Hanya 15 akun non-*spam*mer yang diklasifikasikan sebagai akun *spam*mer. Dan ada 13 akun *spam*mer yang diklasifikasikan bukan pada kelasnya.

Perhitungan Accuration, Precision, Sensitivity, dan Specitivity adalah sebagai berikut :

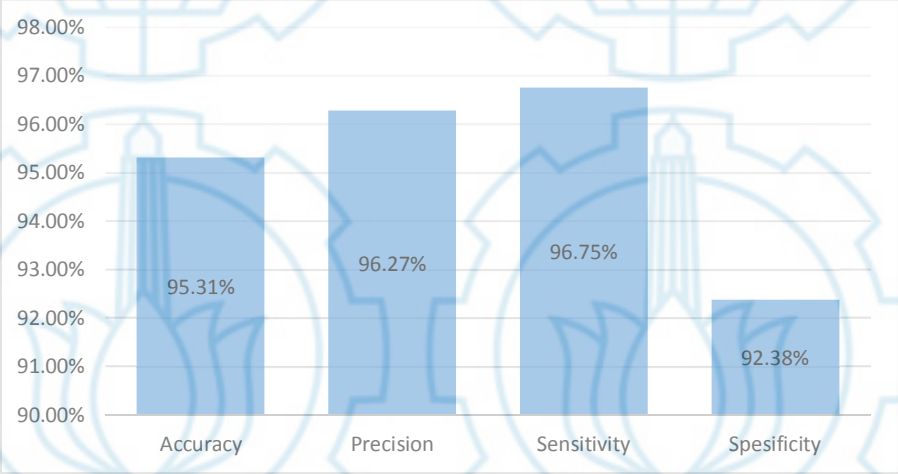
$$\text{Accuracy} = \frac{388+182}{388+182+15+13} \text{ x100\%} = 95{,}31\%$$

$$\text{Precision} = \frac{388}{388+15} \text{ x100\%} = 96{,}27\%$$

$$\text{Sensitivity} = \frac{388}{388+13} \text{ x100\%} = 96{,}75\%$$

$$\text{Specificity} = \frac{182}{182+15} \text{ x 100\%} = 92,\ 38\%$$

Gambar 4.15 menunjukkan hasil dari klasifikasi *spam*mer dan non-*spam*mer menggunakan metode Naïve Bayes berdasarkan kontesn *tweet* jumlah data train 598 data. Hasil accuracy sebesar 95,31%, precision 96,27%, sensitivity 96,75% dan specificity sebesar 92,38%. Hal ini menunjukkan bahwa sistem dapat mengenali akun *spam*mer dengan lebih baik daripada akun non-*spam*mer. Hal ini didukung oleh prosentase sensitivity yang mempunyai nilai yang paling besar. Hasil perhitungan tersebut di atas dapat dibuat grafik seperti di bawah ini :



Gambar 4. 14 Grafik kinerja sistem berdasarkan konten *tweet*

Halaman ini sengaja dikosongkan

BAB 5
KESIMPULAN

Pada bab ini akan diberikan kesimpulan yang dapat diambil oleh penulis selama proses penelitian serta saran-saran ke depan untuk meningkatkan penelitian ini.

5.1 Kesimpulan

Berdasarkan hasil uji coba pada klasifikasi *spam*mer dan non-*spam*mer berdasarkan perilaku pengguna dan berdasarkan konten/isi *tweet*, dapat diambil kesimpulan sebagai berikut:

1. Untuk klasifikasi *spam*mer dan non-*spam*mer berdasarkan profil pengguna *(user-based)* dihasilkan akurasi sebesar 88,235%. Dalam hal ini sistem mengenali kelas non-*spam*mer lebih baik daripada akun *spam*mer.
2. Untuk klasifikasi *spam*mer dan non-*spam*mer berdasarkan konten/isi *tweet (content-based)* dihasilkan akurasi sebesar 95,31%. Dalam hal ini sistem mengenali kelas *spam*mer lebih baik daripada akun non-*spam*mer.
3. *Spam*mer mempunyai rataan limit waktu antar *tweet* lebih kecil daripada non-*spam*mer.
4. Semua *spam*mer adalah akun yang tidak terverifikasi oleh *Twitter*.
5. Dari hasil percobaan klasifikasi spammer berdasarkan profil pengguna *(user-based)* maupun berdasarkan konten/isi *tweet (content-based)* dapat disimpulkan bahwa *non-spammer* tidak mungkin mengirim *tweet* sampah (*spam tweets*), hal ini dapat dilihat dari data kemunculan term terbanyak adalah dari akun spammer, yaitu term “info” dengan kemunculan lebih dari 150 kali.

5.2 Saran

Untuk penelitian selanjutnya dapat dilakukan dengan jumlah akun *spam*mer dan non-*spam*mer dan jumlah *tweet* yang lebih banyak lagi sehingga diperoleh hasil yang lebih baik lagi.

DAFTAR PUSTAKA

[1] "https://www.alexa.com/topsites." .

[2] G. Magno and T. Rodrigues, "Detecting Spammers on Twitter."

[3] A. H. Wang, "DON ' T FOLLOW ME : SPAM DETECTION IN TWITTER," vol. 2010, 2010.

[4] "https://help.twitter.com/en/rules-and-policies/twitter-rules." .

[5] M. McCord and M. Chuah, "Spam Detection on Twitter Using Traditional Classifiers," *Auton. Trust. Comput. - 8th Int. Conf. (ATC 2011)*, vol. 6906 LNCS, pp. 175–186, 2011.

[6] C. Meda, F. Bisio, P. Gastaldo, and R. Zunino, "A machine learning approach for Twitter spammers detection," *Proc. - Int. Carnahan Conf. Secur. Technol.*, vol. 2014–Octob, no. October, 2014.

[7] C. J. Mantas and J. Abellán, "Expert Systems with Applications Credal-C4 . 5 : Decision tree based on imprecise probabilities to classify noisy data," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4625–4637, 2014.

[8] P. G. Metsis V, Androutsopolous I, "Spam Filtering with Naive Bayes Which Naive Bayes.pdf." 2006.

[9] C. Chen *et al.*, "A Performance Evaluation of Machine Learning-Based Streaming Spam Tweets Detection," *IEEE Trans. Comput. Soc. Syst.*, vol. 2, no. 3, pp. 65–76, 2016.

[10] AMIT ANAND AMLESHWARAM, *SPAMMER DETECTION ON ONLINE SOCIAL NETWORKS*, no. December. 2012.

[11] "Https://dev.twitter.com/docs/api/streaming." [Online]. Available: https://dev.twitter.com/docs/api/streaming.

[12] W. Hua and Y. Zhang, "Threshold and associative based classification for social spam profile detection on twitter," *Proc. - 2013 9th Int. Conf.*

Semant. Knowl. Grids, SKG 2013, pp. 113–120, 2013.

[13] M. Mateen, M. Aleem, M. A. Iqbal, and M. A. Islam, "A Hybrid Approach for Spam Detection for Twitter," *14th Int. Bhurban Conf. Appl. Sci. Technol.*, pp. 466–471, 2017.

[14] T. Finin and B. Tseng, "Why we Twitter : Understanding microblogging usage and communities Why We Twitter : Understanding Microblogging," no. January, 2007.

[15] A. Gupta, "Improving spam detection in online social networks," *Cogn. Comput. Inf. Process. (CCIP), 2015 Int. Conf.*, pp. 1–6, 2015.

[16] "REST API resource, https://dev.twitter.com/docs/api." .

[17] E. C. Sammut and G. I. Webb, "Encyclopedia of Machine Learning."

[18] I. S. Ahmad Basuki, *decision Tree*, vol. 46, no. 7. 2004.

[19] M. R. Faisal, "Seri Belajar Data Science: Klasifikasi dengan Bahasa Pemrograman R," *Indones. Net Dev. Community*, pp. 1–50, 2016.

[20] "https.//help.twitter.com/en/managing-your- account/about-twitter-verified-accounts."

[21] S. Karyadi and H. Yasin, "Analisis Kecenderungan Informasi Dengan Menggunakan Metode Text Mining," *Junal Gaussian*, vol. 5, no. TEXY MINING, pp. 763–770, 2016.

LAMPIRAN 1. Data akun *spam*mer dan non-*spam*mer

No	akun	label
1	Allendiar	Non*spam*mer
2	MittaDwinda	Non*spam*mer
3	detikcom	Non*spam*mer
4	Nessyalvioriza	Non*spam*mer
5	parentingINA	Non*spam*mer
6	MRanoTryAstra	Non*spam*mer
7	e100ss	Non*spam*mer
8	Aulyadwiw	Non*spam*mer
9	Anggrainims	Non*spam*mer
10	republikaonline	Non*spam*mer
11	kemendag	Non*spam*mer
12	gilangmsp	Non*spam*mer
13	TelkomPromo	Non*spam*mer
14	kaskus	Non*spam*mer
15	infoBMKG	Non*spam*mer
16	KAI121	Non*spam*mer
17	lafatah	Non*spam*mer
18	Fathyanurul	Non*spam*mer
19	Shaunshata	Non*spam*mer
20	JogjaUpdate	Non*spam*mer
21	shellanurandika	Non*spam*mer
22	tentrioktaviani	Non*spam*mer
23	TommySetyono	Non*spam*mer
24	BukuERLANGGA	Non*spam*mer
25	VanoDaniel	Non*spam*mer
26	zmachmobile	Non*spam*mer
27	zanukoston	Non*spam*mer
28	tribunjogja	Non*spam*mer
29	IndosatCare	Non*spam*mer
30	triindonesia	Non*spam*mer
31	ainunyasmin	Non*spam*mer
32	gitalistyaa	Non*spam*mer
33	galaratama	Non*spam*mer
34	Itothagam	Non*spam*mer
35	juliusGdimas	Non*spam*mer
36	ilmalana	Non*spam*mer
37	honeyqisthi	Non*spam*mer
38	liputan6dotcom	Non*spam*mer
39	IndonesiaGaruda	Non*spam*mer

40	TRANS7	Non*spam*mer
41	TipsBizOnline	Non*spam*mer
42	TheComment_NET	Non*spam*mer
43	ITS_campus	Non*spam*mer
44	kompascom	Non*spam*mer
45	infosurabaya	Non*spam*mer
46	kemendag	Non*spam*mer
47	JogjaUpdate	Non*spam*mer
48	Sahadhewa	Non*spam*mer
49	aagym	Non*spam*mer
50	Azizah_Rhma222	*Spam*mer
51	rizma_rohima1	*Spam*mer
52	sonialunna88	*Spam*mer
53	Veronicawrlta	*Spam*mer
54	Vitaplankton	*Spam*mer
55	Wolf_X9	*Spam*mer
56	Ambimannyu	*Spam*mer
57	YulliYr	*Spam*mer
58	PanggilWanda1	*Spam*mer
59	Cristiisarah	*Spam*mer
60	CoecuCahyati	*Spam*mer
61	DelinaFauziah	*Spam*mer
62	EgaAuliania	*Spam*mer
63	TinnaSintia	*Spam*mer
64	Goodgrow_Ks	*Spam*mer
65	HariTerlanjur	*Spam*mer
66	meida_adinda	*Spam*mer
67	Mmelani2	*Spam*mer
68	Mutilestarii	*Spam*mer
69	NidaSri_Andini	*Spam*mer
70	NadiraPermana	*Spam*mer
71	NaddaYullia	*Spam*mer
72	alya_putri25	*Spam*mer
73	AlinaLiania	*Spam*mer
74	DiyaanaAullia	*Spam*mer
75	bei70xxz	*Spam*mer
76	CinAuliania	*Spam*mer
77	Fany_Herliani	*Spam*mer
78	NaimaHastari	*Spam*mer
79	IcaHans	*Spam*mer
80	sayna_az	*Spam*mer
81	CintaCllara	*Spam*mer

82	meyrenata27	*Spam*mer
83	anakartika_17	*Spam*mer
84	KuisBerhadiahID	*Spam*mer
85	KabarSehatID	*Spam*mer
86	AuraMaheera	*Spam*mer
87	cintaliinda	*Spam*mer
88	blogger_jateng	*Spam*mer
89	cintaanndiina	*Spam*mer
90	JoeliaMarpaung	*Spam*mer
91	EsaAnatasya	*Spam*mer
92	cepatalami	*Spam*mer
93	NadiraPermana	*Spam*mer
94	RefinaAulliana	*Spam*mer
95	kiran_chandra14	*Spam*mer
96	EvanaPaulina	*Spam*mer
97	Emamatina	*Spam*mer
98	AmandaSenjaya	*Spam*mer
99	PanggilIchaaja	*Spam*mer
100	DiyaanaAullia	*Spam*mer

LAMPIRAN 2. Package *twitteR* yang digunakan dalam penelitian

Fungsi	Atribut	Keterangan	Tipe data
getUser	description	Deskripsi pengguna	Char
getUser	statusesCount	Jumlah status pengguna	Numerik
getUser	followersCount	Jumlah followers pengguna	Numerik
getUser	favoritesCount	Jumlah status yang difavoritkan	Numerik
getUser	friendsCount	Jumlah following pengguna	Numerik
getUser	URL	URL yang terkait dengan pengguna	Char
getUser	name	Nama akun pengguna	Char
getUser	created	Waktu akun pengguna dibuat	Datetime
getUser	screenname	Screen name pengguna	Char
getUser	location	Lokasi pengguna	Char
getUser	id	ID pengguna	Char
getUser	listedCount	Berapa kali pengguna muncul dalam daftar umum	Char
getUser	followRequestCount	Jumlah pengguna lain yang mem-follow	Numerik
userTimeline	text	Status pengguna	Char
userTimeline	favorite	Apakah status ini menjadi favorit	Boolean
userTimeline	favoritedCount	Jumlah favorit terhadap status tersebut	Numerik
userTimeline	replyToSN	Screenname pengguna lain yang membalas status	Char
userTimeline	created	Waktu status dibuat	Datetime
userTimeline	truncated	Apakah status ini terpotong	Char
userTimeline	replyToUID	ID pengguna lain yang membalas status	Char
userTimeline	id	ID status	Char
userTimeline	statusSource	Perantara sumber pengguna untuk tweet	Char
userTimeline	screenname	Screenname pengguna yang memasang status	Char
userTimeline	retweetCount	Berapa kali status tersebut di retweet	Numerik
userTimeline	isRetweet	True jika merupakan status yang me-retweet	Boolean
userTimeline	retweeted	True jika merupakan status yang di-retweet	Boolean
userTimeline	longitude	Koordinat garis bujur dari status yang di posting	Char

userTimeline	latitude	Koordinat garis lintang dari status yang di posting	Char
userTimeline	replyToSID	ID pengguna lain yang membalas status	Char

Halaman ini sengaja dikosongkan

BIODATA

Yuli Fitriani, anak ke-4 dari 4 bersaudara dari pasangan M Takim Santoy dan Hj Suwarti, lahir di Sidoarjo, 12 Juli 1983. Penulis menyelesaikan pendidikan Sarjana S1 pada tahun 2010 pada bidang Telekomunikasi Multimedia di Institut Teknologi Sepuluh Nopember ITS. Penulis kemudian melanjutkan pendidikan program Pasca Sarjana S2 pada bidang Jaringan Cerdas Multimedia di Teknik Elektro ITS pada tahun 2017.

Penulis aktif di Masjid Nasional Al Akbar Surabaya sebagai staf IT, selain itu penulis juga pernah mengajar di STAI Masjid Nasional Al Akbar Surabaya sebagai staf pengajar mata kuliah Web Design dan Design Graphic. Penulis dapat dihubungi melalui email yulifit@gmail.com.