Jurnal Ilmu Komputer Agri-Informatika

Tersedia secara online di:
http://journal.ipb.ac.id/index.php.jika

Volume 2 Nomor 1 halaman 1-8
ISSN: 2089-6026

Pengukuran Kinerja *Spam Filter* Menggunakan *Graham's Naïve Bayes Classifier*

Measuring Performance of Spam Filter using Graham's Naive Bayes Classifier

JULIO ADISANTOSO*, WILDAN RAHMAN

Abstrak

Email *spam* telah menjadi masalah utama bagi pengguna dan penyedia jasa Internet. Pendekatan *heuristic* telah dilakukan untuk menyaring *spam* seperti *black-listing* atau *rule-based filtering*, namun hasilnya kurang memuaskan sehingga pendekatan berbasis konten (*content-based filtering*) menggunakan pengklasifikasi naïve Bayes lebih banyak digunakan saat ini. Penelitian ini bertujuan membandingkan pengklasifikasi naïve Bayes multinomial yang menggunakan atribut *boolean* dengan versi Graham, dan juga membandingkan kinerja dari dua metode untuk data latih, yaitu *train-everything* (TEFT) dan *train-on-error* (TOE). Hasil evaluasi menunjukkan bahwa naïve Bayes multinomial memiliki kinerja lebih baik dibanding versi Graham. Di samping itu, metode data latih menggunakan TEFT dapat meningkatkan akurasi model klasifikasi dibanding metode TOE.

Kata kunci: filter *spam*, naïve Bayes, metode *training*

Abstract

Spam email has become a major problem for Internet users and providers. After several failed attempt to filter spam based on heuristic approach such as black-listing or rule-based filtering, content-based filtering using naïve Bayes classifier has become the standard for spam filtering today. However, the naive Bayes classifier exists in different forms. This research aims to compare two different forms of naïve Bayes which are multinomial naive Bayes using boolean attribute and Graham version of naïve Bayes which is popular among several commercial and open source spam filter applications. This research also compares performace of two different methods for data trainings which are train-everything (TEFT) and train-on-error (TOE). Finally, this research attempts to identify several hard-to-classify emails. The evaluation result showed that multinomial naïve Bayes had better performance compared to Graham naïve Bayes. The result also showed that TEFT successfully outperforms TOE in term of accuracy.

Keywords: spam filter, naïve Bayes, training method

PENDAHULUAN

Pemanfaatan teknologi jaringan Internet yang semakin meningkat intensitasnya dewasa ini berdampak besar pada metode pengiriman surat. Jalur fisik yang semula menjadi pilihan semakin ditinggalkan dan digantikan oleh jalur pengiriman elektronik dalam bentuk *electronic mail* atau biasa disebut dengan *email*. Berbagai macam keunggulan yang dimiliki oleh *email* ternyata banyak disalahgunakan untuk mengirimkan pesan berbau komersial secara massal.

Spam adalah pesan atau *email* yang "tidak diinginkan" oleh penerimanya dan dikirimkan secara massal. Makna "tidak diinginkan" di sini memiliki arti pihak pengirim tidak mendapatkan izin untuk mengirimkan pesan tersebut dari pihak penerima. Makna "dikirimkan secara massal" berarti pesan tersebut merupakan bagian dari sekumpulan pesan yang memiliki isi yang sama atau sejenis dan dikirimkan sekaligus dalam jumlah besar (Spamhaus 2009).

Berbagai upaya telah dilakukan untuk mengatasi permasalahan *spam*. Pada mulanya proses penyaringan *email spam* dilakukan dengan pendekatan *rule-based*. Email dikategorikan sebagai *spam* menurut aturan-aturan tertentu seperti kemunculan kata, alamat pengirim, dan struktur *header*. Pendekatan ini dalam praktiknya kurang efektif dan memiliki tingkat *false positive* yang tinggi. Selain *rule-based*, metode *spam-filtering* yang banyak digunakan di masa lalu adalah *server blacklist* dan *signature-based filtering* (Graham 2003).

Pendekatan berbasis *content statistic* (menggunakan metode Naïve Bayes) untuk menyaring pesan *spam* pertama kali diteliti oleh Pantel dan Lin (1998) dan berhasil mencapai tingkat akurasi 92% dengan tingkat *false positive* sebesar 1.16%. Teknik serupa juga digunakan oleh Sahami *et al*. (1998) meskipun kinerjanya tidak setinggi filter *spam* yang dirancang oleh Pantel dan Lin (1998).

Graham (2002) membahas teknik *spam-filtering* menggunakan metode pengklasifikasi naïve Bayes (NB) dengan pendekatan yang cukup berbeda jika dibandingkan dengan metode pengklasifikasi naïve Bayes pada umumnya. Metode ini diklaim berhasil mencapai tingkat akurasi sebesar 99.95% dengan *false positive* sebesar 0.05%. Kinerja yang cukup tinggi ini membuat metode *content-based filtering* semakin banyak digunakan dalam aplikasi filter *spam* (Yerazunis 2004).

Penelitian ini menguji dua model dari teknik klasifikasi NB, yaitu NB multinomial dengan atribut *boolean* dan NB Graham. Penelitian ini juga menguji dua metode *training* yang digunakan pada filter *spam*. Lebih lanjut, penelitian ini membahas komponen pendukung yang digunakan dalam pembuatan filter *spam* seperti teknik pemrosesan *email*, pemilihan fitur, dan tokenisasi.

METODE

Penelitian ini terdiri atas empat tahap, yaitu pengumpulan data, pengujian metode *training*, pengujian metode klasifikasi, dan analisis kesalahan klasifikasi. Untuk metode *training*, terdapat dua macam metode yang diuji, yaitu *training everything* (TEFT) dan *training on error* (TOE). Masing-masing metode *training* diduga memiliki kinerja yang berbeda sehingga dilakukan pengujian untuk menentukan metode *training* mana yang memiliki kinerja yang lebih baik.

Pada metode TEFT, seluruh *email* yang masuk akan dilatih tanpa memperhatikan kebenaran hasil klasifikasinya. Kelebihan dari metode ini adalah sekumpulan data dalam filter *spam* yang akan terus menyesuaikan nilainya sesuai dengan *email* yang diterimanya. Sebagai contoh, jika pengguna berlangganan *mailing-list* tertentu, filter akan segera mengenali token-token di dalamnya sebagai bagian dari kelas *ham*. Sedangkan pada metode TOE, *email* hanya akan dimasukkan ke dalam proses *training* jika terjadi kesalahan klasifikasi. Kelebihan metode TOE idalah proses *training* hanya dilakukan seperlunya sehingga menghemat sumber daya, seperti proses *disk-writing* yang lambat. Metode TOE juga menyimpan lebih sedikit token sehingga dapat menghemat ruang penyimpanan.

Metode klasifikasi yang diuji terdiri atas dua model NB, yaitu Bayes multinomial dengan atribut *boolean* dan metode Bayes Graham. Bayes multinomial dengan atribut *boolean* hampir sama dengan Bayes multinomial yang menggunakan atribut frekuensi token atau *term frequency* (**tf**), termasuk juga proses pendugaan nilai peluang suatu token ke-$\boldsymbol{i}$ ($\boldsymbol{t_i}$) dalam kelas *spam* $\boldsymbol{S}$, yaitu $\mathbf{P}(\boldsymbol{t_i}|\boldsymbol{S})$. Perbedaannya terletak pada atribut yang digunakan, yaitu bernilai *boolean*. Pada atribut *boolean*, token yang muncul berulang kali tetap dianggap sebagai satu kemunculan untuk setiap dokumennya. Schneider (2004) dalam penelitiannya menunjukkan bahwa Bayes multinomial akan memiliki kinerja yang lebih baik jika atribut frekuensi token digantikan dengan atribut *boolean*.

Graham (2002) menggunakan pendekatan yang berbeda dalam mengimplementasikan naïve Bayes. Jika pada metode sebelumnya digunakan pendugaan nilai $\mathbf{P}(\boldsymbol{t_i}|\boldsymbol{S})$ untuk

mendapatkan peluang suatu *email* masuk ke dalam kelas *spam*, metode Bayes Graham menggunakan pendugaan nilai $\mathbf{P}(\boldsymbol{S}|\boldsymbol{t_i})$ untuk menghitung peluang suatu *email* masuk ke dalam kategori *spam* jika diketahui *email* tersebut mengandung token $\boldsymbol{t_i}$, yaitu:

$$P(S|t_i) = \frac{\frac{f_{is}}{n_s}}{\frac{f_{is}}{n_s} + \frac{f_{ih}}{n_h}} \quad (1)$$

dengan $\boldsymbol{f_{is}}$ dan $\boldsymbol{f_{ih}}$ berturut-turut adalah banyaknya *email* pada kelas *spam* dan *ham* yang mengandung token ke-$\boldsymbol{i}$, serta $\boldsymbol{n_s}$ dan $\boldsymbol{n_h}$ berturut-turut adalah jumlah pesan yang tedapat pada kelas *spam* dan *ham*. Semakin sering suatu token muncul di kelas *spam*, nilai peluangnya akan semakin mendekati satu (Crossan 2009).

Untuk menghitung peluang suatu *email* masuk ke dalam kelas *spam*, metode Bayes Graham hanya menggunakan 15 token yang paling signifikan. Seberapa signifikan suatu token dalam menentukan hasil klasifikasi ditentukan dengan melihat selisih nilai $\mathbf{P}(\boldsymbol{S}|\boldsymbol{t_i})$ terhadap nilai peluang netral 0.5. Kelima belas token yang paling signifikan tersebut digunakan untuk menghitung peluang suatu *email* masuk ke dalam kelas *spam* dengan Persamaan 2.

$$P(S|t_1,\ldots,t_g) = \frac{\prod_{i=1}^{15} P(S|t_i)}{\sum_{C' \epsilon (S,H)} \prod_{i=1}^{15} P(C'|t_i)} \quad (2)$$

Pesan akan dikategorikan sebagai *spam* jika nilai pada Persamaan 1 bernilai lebih besar dari 0.9.

Pengumpulan Data

Tahap penelitian yang pertama adalah tahap pengumpulan data. Data yang digunakan sebagai data uji adalah korpus *email* dalam format asli, yang masih memiliki bagian *header* dan *body*. Data ini berisi campuran pesan yang sudah diberi label kelas, yaitu '*ham'* dan '*spam'* sesuai dengan kelasnya. Proses pemberian label kelas ini dilakukan secara manual.

Pengujian Metode *Training* dan Metode Klasifikasi

Pengujian dua metode *training* (TEFT dan TOE) dilakukan dengan cara mengukur akurasi kedua metode tersebut saat dipasangkan dengan metode Graham. Untuk menguji metode klasifikasi Bayes multinomial yang menggunakan atribut *boolean* digunakan metode yang sama dengan metode evaluasi pada penelitian Yerazunis (2004), yaitu:

1 Data uji berupa korpus *email* yang sudah diklasifikasikan ke dalam dua kelas, yaitu '*ham'* dan '*spam'* disediakan. Setiap *email* diberi label sesuai dengan kelasnya. Data yang sudah diberi label tersebut kemudian digabungkan.
2 Data uji kemudian diacak sebanyak sepuluh kali. Setiap hasil pengacakan dicatat urutan pembacaannya sehingga seluruh metode yang diuji dievaluasi menggunakan acakan dan urutan pembacaan data yang sama.
3 Untuk setiap acakan, sebanyak *N* data yang nantinya digunakan sebagai data *testing* akhir diambil.
4 Proses pengujian dilakukan sebanyak jumlah acakan, yaitu sepuluh kali pengujian.
5 Langkah 1 sampai dengan 4 akan menghasilkan data awal hasil pengujian berupa jumlah kesalahan klasifikasi dari 10 kali *N* data uji.

Data awal hasil pengujian diolah lagi untuk mendapatkan tingkat akurasi hasil prediksi berupa jumlah *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN) seperti yang dapat dilihat pada Tabel 1.Hasil positif merujuk pada *email* yang diprediksikan masuk ke dalam kategori *spam* dan hasil negatif merujuk pada *email* yang diprediksikan masuk ke dalam kategori *ham* oleh filter. Kinerja masing-masing metode dievaluasi dengan melihat nilai dari *spam recall* dan *ham recall*. *Spam recall* adalah proporsi dari *email spam* yang berhasil diblok oleh filter, sedangkan *ham recall* menunjukkan proporsi dari *email ham* yang diloloskan oleh filter (Metsis *et al*. 2006).

Tabel 1 Tabel kontingensi kelas hasil prediksi dan kelas sebenarnya

		Kelas Prediksi	
		Spam	*Ham*
Kelas Sebenarnya	*Spam*	TP	FN
	Ham	FP	TN

Analisis Kesalahan Klasifikasi

Setelah pengujian untuk metode *training* dan metode klasifikasi selesai dilakukan, penelitian selanjutnya berfokus pada analisis kesalahan klasifikasi. Pesan-pesan yang gagal diklasifikasikan ke dalam kelas yang benar, akan diteliti lebih lanjut untuk mencari penyebab kegagalan klasifikasi.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Korpus yang digunakan pada penelitian ini adalah *public email corpus* yang disediakan oleh SpamAssasin yang diunduh dari alamat http://www.spamassassin.org/publiccorpus/ dengan kode prefiks 20030228. Korpus ini terdiri atas 6047 pesan *email* yang sudah diklasifikasikan sebelumnya secara manual. Perbandingan jumlah *spam* dan *ham* untuk data uji pada masing-masing acakan dapat dilihat pada Tabel 2.

Tabel 2 Proporsi pesan *spam* untuk masing-masing acakan pengujian

Pengacakan ke	Jumlah *ham*	Jumlah *spam*	Persen *spam*
1	529	221	29.46
2	512	238	31.73
3	514	236	31.46
4	528	222	29.60
5	518	232	30.93
6	484	266	35.46
7	518	232	30.93
8	508	242	32.26
9	529	221	29.46
10	511	239	31.86
Jumlah	5151	2349	31.32

Pemrosesan Dokumen

Untuk fase latihan dan fase pengujian, setiap *email* diproses dengan teknik yang sama. Pemrosesan yang dilakukan terdiri atas empat tahap yaitu dekomposisi struktur *email*, pemilihan atribut, penyeragaman sistem karakter, dan tokenisasi. Secara garis besar, tahapan dekomposisi *email* yang dilakukan sebagai berikut:

1 Email dipecah ke dalam dua bagian utama, yaitu *header* dan *body*.
2 Komponen *header* dipecah lagi menjadi komponen-komponen yang lebih kecil sesuai dengan informasi yang dikandungnya.
3 Untuk komponen *body*, pesan yang terdiri atas beberapa bagian akan digabungkan menjadi satu. Jika pada *email* terdapat *attachment*, hanya informasi nama dan jenis *file* yang disertakan.

Setelah *email* dipecah menjadi komponen-komponen yang lebih kecil, tahapan selanjutnya adalah pemilihan komponen yang akan disertakan ke dalam proses klasifikasi. Tahapan ini berlaku terutama untuk bagian *header* dari *email*.

Tidak semua komponen dari *header* dimasukkan ke dalam klasifikasi karena terdapat beberapa informasi pada *header* yang telah mengalami kerusakan ataupun telah diubah sebelumnya oleh pihak SpamAssasin sebagai penyedia data. Selain itu, terdapat komponen *header* yang hanya muncul di sebagian kecil dokumen. Komponen-komponen tersebut adalah informasi tambahan yang biasanya disertakan oleh klien *email* atau *Mail Transfer Agent* yang dilalui oleh *email* sebelum sampai ke tujuan (Tabel 3).

Tabel 3 Komponen *header* yang disertakan dalam proses klasifikasi

Nama	Keterangan
Subject	Subjek dari pesan.
Sender	Nama dan alamat pengirim pesan.
return-path	Alamat pengembalian pesan jika terjadi *bouncing* (kondisi dimana alamat penerima tidak ditemukan).
x-mailer	Aplikasi yang digunakan oleh pengguna untuk mengirimkan pesan.
reply-to	Alamat yang digunakan untuk membalas pesan.
content-transfer-encoding	Metode *content transfer encoding* yang digunakan jika ada.

Hasil Pengujian Metode *Training*

Pengujian metode *training* dilakukan dengan cara memasangkan kedua metode tersebut pada filter *spam* yang menggunakan metode klasifikasi Bayes Graham. Pada metode TEFT, seluruh *email* yang dibaca akan dimasukkan ke dalam kelas yang benar setelah hasil dari klasifikasi diperoleh. Proses *training* ini dilakukan tanpa memperhatikan hasil klasifikasinya benar atau salah. Pada metode TOE, proses *training* hanya akan dilakukan jika terjadi kesalahan klasifikasi.

Jumlah FP dan FN per 7500 kali pengujian beserta *ham recall* dan *spam recall* dapat dilihat pada Tabel 4.

Tabel 4 Hasil pengujian metode *training* menggunakan teknik klasifikasi Graham

	TEFT	TOE
False Positive	74	475
False Negative	50	67
Spam Recall	0.9786	0.9714
Ham Recall	0.9856	0.9079

Hasil pengujian menggunakan metode Bayes Graham menunjukkan bahwa metode *training* TEFT memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan metode *training* TOE. Perbedaan akurasi ini disebabkan oleh lebih banyaknya proses *training* yang dilakukan oleh metode TEFT dibandingkan dengan metode TOE. Proses *training* yang lebih banyak membuat metode TEFT menyimpan informasi yang lebih akurat mengenai karakteristik token-token dari kelas *spam* maupun *ham* dalam data hasil *training*. Walaupun perbedaan nilai *spam recall* hanya sebesar 0.0072, namun perbedaan nilai *ham recall* antara kedua metode *training* tersebut cukup tinggi, yaitu 0.0777.

Nilai *ham recall* berhubungan dengan tingkat *false positive*. Pada filter *spam*, *overhead* dari *false positive* lebih tinggi dibandingkan dengan *false negative*. Berdasarkan hal ini, perbedaan tingkat akurasi ini cukup signifikan untuk dipertimbangkan.

Karena proses *training* dilakukan untuk seluruh *email* yang masuk, TEFT membutuhkan waktu pengujian lebih lama dibandingkan dengan TOE. Hasil pengujian

menunjukkan TEFT menghabiskan waktu sekitar 18% lebih lama dibandingkan dengan TOE. Dengan demikian, meskipun TOE memiliki tingkat akurasi yang lebih rendah dibandingkan dengan TEFT, waktu pemrosesan yang dilakukan oleh TOE lebih singkat.

Perbedaan waktu antara kedua metode *training* ini dapat dibandingkan dengan memperhatikan perbedaan nilai *spam recall* dan *ham recall* untuk setiap tambahan waktu proses. Untuk *spam recall*, peningkatan kinerja per satuan waktu (***GS***) dapat dihitung dengan Persamaan 3.

$$GS = \left| \frac{SR_{TOE} - SR_{TEFT}}{DW} \right| \quad (3)$$

dengan ***SR*** adalah nilai *spam recall* dan ***DW*** adalah persentase perbedaan waktu yang dihabiskan oleh kedua metode *training*. Dengan cara yang sama, peningkatan *ham recall* (***GH***) untuk masing-masing metode *training* dapat dihitung dengan Persamaan 4.

$$GH = \left| \frac{HR_{TOE} - HR_{TEFT}}{DW} \right| \quad (4)$$

dengan ***HR*** adalah nilai *ham recall* untuk masing-masing metode *training*. Dibandingkan dengan metode TOE, metode TEFT dapat meningkatkan akurasi namun prosesnya lebih lambat. Dengan perhitungan pada Persamaan 3 dan 4, didapat bahwa penggunaan metode TEFT dibandingkan dengan TOE akan meningkatkan *spam recall* sebesar 0.04% untuk setiap 1% penambahan waktunya. Sementara untuk *ham recall*, peningkatan akurasi yang didapatkan adalah sebesar 0.43% untuk setiap 1% penambahan waktunya.

Hasil Pengujian Metode Klasifikasi

Proses pengujian metode klasifikasi dilakukan dengan menggunakan metode *training* TEFT. Jumlah FP dan FN per 7500 kali pengujian beserta nilai *ham recall* dan *spam recall* dapat dilihat pada Tabel 5.

Hasil pengujian menggunakan mode *training* TEFT menunjukkan bahwa metode Bayes Graham memiliki *spam recall* lebih tinggi dibanding metode Bayes multinomial dengan perbedaan nilai sebesar 0.0171. Hasil sebaliknya terlihat pada *ham recall* dimana metode Bayes multinomial memiliki nilai yang lebih tinggi dengan perbedaan nilai sebesar 0.0150.

Tabel 5 Hasil pengujian metode klasifikasi dengan metode *training* TEFT dan TOE

	Metode TEFT		Metode TOE	
	Graham	Multinomial	Graham	Multinomial
False Positive	74	70	475	117
False Negative	50	67	67	55
Spam Recall	0.9786	0.9615	0.9714	0.9765
Ham Recall	0.9714	0.9864	0.9079	0.9773

Pengujian dengan metode *training* TOE menunjukkan bahwa metode klasifikasi Bayes multinomial memiliki *spam recall* dan *ham recall* yang lebih tinggi dibandingkan dengan metode Bayes Graham dengan perbedaan masing-masing sebesar 0.0051 dan 0.0694.

Analisis Kesalahan Pengenalan Ham (*False Positive*)

Meskipun dalam proses klasifikasinya metode Bayes Graham mengalikan jumlah kemunculan token pada kelas *ham* dengan faktor bernilai 2, ternyata tingkat *ham recall* masih lebih rendah dibandingkan dengan *ham recall* dari metode Bayes multinomial. Lebih rendahnya *ham recall* dari metode Bayes Graham disebabkan oleh pemberian nilai 0.99 untuk token yang hanya pernah muncul di kelas *spam*. Dalam menentukan hasil klasifikasi, metode

Bayes Graham hanya menggunakan 15 token yang paling signifikan. Jika *email* dari kelas *ham* mengandung token-token yang hanya muncul di kelas *spam*, proses klasifikasi akan didominasi oleh token-token *spam* karena token dengan peluang 0.99 memiliki selisih yang tinggi dari peluang netral 0.5.

Adapun karakteristik *email* yang menyebabkan *false positive* antara lain ialah:

Email *ham* yang mengandung tag HTML

Karakteristik seperti ini banyak ditemukan pada *email* yang berjenis *newsletter*. Karena data uji yang digunakan tidak memiliki *email newsletter* dalam jumlah yang cukup, token-token *ham* pada *email* tersebut tidak memiliki nilai $\mathbf{P}(\boldsymbol{H}|\boldsymbol{t})$ yang signifikan untuk mengimbangi token-token *spam* yang ada. Akibatnya terjadi *false positive* dalam proses *filtering* yang dilakukan.

Newsletter resmi yang memiliki isi bertema promosi

Selain pengaruh tag HTML, *false positive* juga banyak dipengaruhi oleh isi dari *email* itu sendiri. Meskipun *newsletter* dikirimkan dengan persetujuan penerimanya, isi dari *newsletter* seringkali berbau promosi dan menggunakan kata-kata yang digunakan pada *email spam*. Pada kasus seperti ini, baik metode Graham maupun multinomial, keduanya mengalami kesulitan dalam menentukan kelas yang benar.

Email *ham* yang memiliki beberapa format *multipart-alternative*

Email dengan format *multipart-alternative* memungkinkan *email* dikirimkan dalam beberapa versi sekaligus. Sebagai contoh, jika aplikasi klien memiliki kapabilitas untuk membaca dokumen HTML, maka akan ditampilkan versi *email* yang menggunakan tag HTML. Namun jika tidak, akan ditampilkan versi yang hanya menggunakan teks biasa.

Kesalahan lain

Selain beberapa butir yang sudah disebutkan, terdapat beberapa faktor lain yang menyebabkan *false positive* meskipun tidak dalam jumlah banyak. Email yang menggunakan token berhuruf kapital dalam jumlah banyak, seperti *email* berisi peringatan cuaca buruk, sering salah diklasifikasikan sebagai *spam* karena kata-kata dalam huruf kapital banyak ditemukan pada *email spam*. Email yang dikirimkan oleh aplikasi *auto-responder* beberapa kali salah diklasifikasikan sebagai *spam* karena isinya yang pendek dan mengandung kata-kata yang umum ditemukan pada *email spam* seperti '*call*', '*contact'*, dan '*respond'*. Selain itu, *email* pendek yang hanya berisikan URL juga kadang salah diklasifikasikan sebagai *spam* karena URL lebih banyak ditemukan pada *email spam*.

Analisis Kesalahan Pengenalan Spam (*False Negative*)

Hasil pengujian terhadap kedua metode klasifikasi tidak membuahkan kesimpulan mengenai metode mana yang memiliki tingkat *false negative* lebih tinggi. Selanjutnya, karakteristik pesan spam yang berhasil lolos dari proses *filtering* dibahas.

Email *spam* yang kebetulan memiliki isi seperti *ham*.

Pesan-pesan *ham* pada data uji didominasi oleh pesan dari *mailing list* bertemakan teknologi informasi. Pesan-pesan *spam* yang memiliki tema sangat berbeda seperti obat ataupun judi akan mudah dikenali oleh filter *spam*, namun jika *spam* yang dikirim ternyata bertemakan teknologi informasi, kata-kata yang terkandung di dalamnya akan memiliki karakteristik yang mirip dengan mayoritas *email ham*. Akibatnya, filter akan salah mengklasifikasikan *email spam* tersebut sebagai *ham*.

Email yang menggunakan huruf non-latin.

Pemrosesan tokenisasi pada *email* berkarakter latin dan berkarakter non-latin memiliki sedikit perbedaan. Pada *email* berkarakter latin, dengan satu karakter pemisah token saja yaitu spasi, *email* sudah dapat dipecahkan menjadi token-token dengan cukup baik. Pada tulisan dimana karakternya tidak menggunakan spasi sebagai pemisah token, proses tokenisasi biasa tidak akan menghasilkan token-token yang sesuai. Selain itu, *email* dengan huruf non-latin

pada data pengujian jumlahnya sangat sedikit. Kurangnya data latih untuk token-token yang ada menyebabkan token tersebut memiliki nilai peluang yang cenderung netral, yaitu 0.5. Khusus pada metode Graham, batas nilai peluang suatu *email* untuk masuk ke dalam kelas *spam* adalah 0.9 sehingga *email-email* yang dipenuhi dengan token netral akan masuk ke dalam kelas *ham*.

Email yang isinya gagal di-*decode*.

Beberapa *email* yang menggunakan *character-encoding* maupun *content-transfer-encoding* khusus gagal di-*decode*. Hal ini disebabkan oleh kurangnya pustaka pembaca *email* yang digunakan atau kesalahan format pada *email*. Kegagalan proses *decode* menyebabkan isi dari *email* hanya muncul sebagian atau tidak muncul sama sekali sehingga proses klasifikasi didominasi oleh token-token dari *header*.

SIMPULAN

Pengujian menggunakan metode klasifikasi Bayes Graham menunjukkan metode *training* TEFT memiliki akurasi yang lebih tinggi dibandingkan dengan metode TOE, terutama pada *ham recall* yang memiliki perbedaan nilai mencapai 0.0777. Pengujian menggunakan kedua metode *training* menunjukkan metode Bayes multinomial memiliki akurasi yang lebih tinggi dibandingkan dengan metode Bayes Graham, kecuali untuk *spam recall* pada pengujian dengan metode *training* TEFT ketika metode Bayes Graham memiliki nilai *spam recall* yang lebih tinggi.

Kegagalan pengenalan *ham* (*false positive*) disebabkan oleh penggunaan token-token yang umum pada *email spam* di kelas *ham* seperti *email* dengan tag HTML, sedangkan kegagalan pengenalan *spam* (*false negative*) disebabkan oleh isi dari *email spam* yang dikirim kebetulan sama dengan tema dari *email* pengguna.

DAFTAR PUSTAKA

Crossan J. 2009. Naïve Bayes classification in spam filtering.

Graham P. 2002. A plan for spam [Internet]. [diunduh 2009 Des 14]. Tersedia pada: http://paulgraham.com/spam.html.

Graham P. 2003. Stopping spam [Internet]. [diunduh 2009 Des 14]. Tersedia pada: http://paulgraham.com/stopspam.html.

Metsis V, Androutsopoulos I, Paliouras G. 2006. Spam filtering with naïve Bayes-which naïve Bayes? Di dalam: *Third Conference on Email and AntiSpam (CEAS) 2006*; 2006 Jul 27-28; Mountain View, Amerika Serikat. 17:28-69.

Pantel P, Lin D. 1998. SpamCop: a spam classification and organization program. Di dalam: *Proceedings of AAAI-98 on Learning for Text Categorization*; 1998 Jul 26-27; Madison, Amerika Serikat. hlm 95-98.

Sahami M, Dumais S, Heckerman D, Horvitz E. 1998. A Bayes approach to filtering junk e-mail. Di dalam: *Proceedings of AAAI-98 on Learning for Text Categorization*; 1998 Jul 26-27; Madison, Amerika Serikat. hlm 98-105.

Schneider KM. 2004. On word frequency information and negative evidence in naïve Bayes text classification. Di dalam: *Advances in Natural Language Processing*. Berlin (DE): Springer. hlm 474-485.

Spamhaus. 2009. The definition of spam [Internet]. [diunduh 2009 Des 29]. Tersedia pada: http://www.spamhaus.org/definition.html.

Yerazunis WS. 2004. The spam-filtering accuracy plateau at 99.9 percent accuracy and how to get past it. Di dalam: *Proc. MIT Spam Conference 2004*.