ISBN: 979-756-061-6

METODE PENYARINGAN EMAIL YANG TIDAK DIINGINKAN MENGGUNAKAN PENDEKATAN PROBABILISTIK

Miftah Andriansyah, Adang Suhendra
Jl. Margonda Raya No. 100, Depok, Universitas Gunadarma, Jakarta
E-mail: didi@staff.gunadarma.ac.id, adang@staff.gunadarma.ac.id

ABSTRAK

Sekarang ini, email menjadi salah satu entitas yang penting dalam hal komunikasi, baik personal, koorporat, pemerintah dan komunitas lainnya. Meluasnya penggunaan email sebagai media komunikasi mempunyai dampak negative yang bermuara pada bertambahnya beban/biaya ekonomis, diantaranya, masalah keamanan, privasi dan efisiensi, time cost, dan lain lain. Email yang pada awalnya dimaksudkan hanya sebagai alat komunikasi menjadi lebih luas penggunaannya mulai dari aktifitas ekonomi, hingga aktifitas berdampak negative, yang menjadi fokus penulisan ini. Dampak negative yang dihasilkan bisa berupa email yang tidak diinginkan (misal, junk mail atau spam) oleh pengguna baik institusi maupun pribadi hingga email yang mengandung virus, worm atau entitas lainnya yang dapat merugikan pengguna Untuk menangkal dampak-dampak negative (penggunaan email) yang mungkin terjadi diperlukan proses pemilihan/penyaringan terhadap konten yang melekat pada email tersebut.

Ada dua kategori metode penyaringan berdasarkan cara kerjanya, yaitu: statik dan dinamik (aktif). Metode statik banyak digunakan pada software penyaringan email generasi lama. Metode dinamik merupakan paradigma baru yang menggantikan konsep statik, dimana kedua metode tetap menggunakan pinsip matematik, namun yang membedakannya adalah penggunaan pendekatan probabilistik pada metode penyaringan dinamik, dimana suatu email dikategorikan baik atau buruk (tidak diinginkan) terkait dengan kejadian-kejadiannya dimasa lalu, untuk itu, metode dinamik bias disebut metode probabilistik.

Dalam banyak penelitian menyebutkan bahwa metode probabilistik dapat menyaring email yang tidak diinginkan dengan tingkat keakuratan lebih dari 95%, untuk itu penggunaan konsep penyaring email dengan metode probabilistik untuk mengurangi dampak ekonomi negative penggunaan email sebagai elemen aktifitas ekonomi.

Kata kunci: *email, metode probabilistik, Filter Bayes, anti- spam,*

1. PENDAHULUAN

Penggunaan email sebagai salah satu entitas penting dalam komunikasi digital, terutama Internet, mendorong secara langsung maupun tidak langsung bagi banyak pihak, terutama yang tidak bertanggung jawab untuk mengambil keuntungan yang tidak wajar atau hanya sekedar untuk kepentingan pribadi semata. Penyebaran *email* yang tidak dinginkan, yang disebut sebagai *spam,* pada lalu-lintas Internet bisa berdampak pada rendahnya efisiensi dan menrurunnya produktifitas pekerjaan pengguna yang mengkases institusi pemerintah, akademik dan terutama pada organisasi bisnis. Yang sering terjadi yaitu, banyak waktu kerja yang terbuang hanya untuk menghapus, misal *spam*, yang dalam hitungan hari, minggu, bulan dan tahunnya. Pertanyaan yang akan dialamatkan oleh banyak institusi adalah, "Berapa banyak biaya yang harus ditanggung institusi saya ?" untuk menghilangkan atau mengahapus email yang tidak dinginkan! Inefisiensi tersebut merupakan dampak yang muncul sebagai akibat dari para pengguna yang kehilangan waktu (secara sadar maupun tidak sadar) kerja optimal untuk menyelesaikan pekerjaannya pada saat iyang bersamaan dengan kegiatan pengahapusan email.

Ada beberapa metode untuk menanggulangi masalah tersebut, yang kami kategorikan menjadi dua kategori, yaitu metode static dan metode dinamik. Metode statik adalah meotde yang selam ini telah digunakan pada banyak software penyaringan email, namun pada implementasi, metod tersebut mempunyai kelemahan yaitu tidak dapat menyaring email yang telah dimodifikasi/dimanipulasi oleh para *spammer* (pihak yang mengirim *spam*) yang semakin hari semakin "kreatif".

Metode dinamik, adalah metode yang menghitung probabilitas email terbaru yang masuk terhadap email yang telah masuk sebelumnya, yang disimpan di database.

Dalam banyak kasus, keuntungan utama dari metode ini adalah tingkat keakuratannya yang tinggi dalam mendeteksi dan "membunuh" *spam* dengan nilai yang signifikan hingga 99,9% [1].

2. METODOLOGI

Metodologi yang digunakan dalam penulisan ini adalah metode probabilistik dengan mengadopsi metode Bayes yang menjelaskan bahwa perhitungan probabilistik suatu kejadian sekarang menpunyai kaitan dengan kejadian sebelumnya. Pengambilan keputusan suatu kejadian adalah

 ISBN: 979-756-061-6

benar/ salah bergantung pada kejadian-kejadian sebelumnya. Praktiknya, suatu email dinyatakan sebagai suatu spam/ham ditentukan probabilitasnya berdasarkan email tersebut berada dalam database spam atau database ham. Tingkat toleransi suatu email sebagai spam ditentukan oleh desainernya, semakin tinggi tingkat toleransi mengkategorikan sebagai spam, maka semakin tinggi pula keakurasiannnya dalam memblok email tersebut.

3. TEKNIK PENYARINGAN PROBABILSITIK

Dalam teknik probabilistiks, metode matematik yang digunakan adalah metode Bayesian yang bersandar pada teori Bayes, yaitu menghitung besarnya probabilitas email yang terbaru dibandingkan dengan probabiltas email yang ada/masuk sebelumnya yang tersimpan dalam database email.

3.1 Teori Bayes

Teori Bayesian diadopsi dari nama penemunya yaitu Thomas Bayes sekitar tahun 1950, yang sering ditemukan pada studi-studi ilmu statistika yang berbasis pada teorema atau aturan Bayes.

Teori Bayesian adalah sebuah teori kondisi probabilitas yang memperhitungkan probabilitas suatu kejadian (hipotesis) bergantung pada kejadian lain (bukti). Pada dasarnya, teorema tersebut mengatakan bahwa kejadian di masa depan dapat diprediksi dengan syarat kejadian sebelumnya telah terjadi.

Statement teori bayes:

$$P(A|B)P(B) = P(A,B) = P(B|A)P(A)$$

dimana *P(A/B)* adalah probabilitas gabungan kejadian *A* dan *B*. Membagi kedua sisi dengan *P(B)* , didapat:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Secara umum teorema Bayes dapat dituliskan dalam bentuk:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)},$$

Jika {*Ai*} membentuk partisi dari ruang kejadian, untuk setiap *Ai* dalam partisi.

Variasi lain dari teori Bayes untuk anti*spam* adalah *Naïve Bayesian Filtering* (Bayesian Murni).

3.2 Filtering *Anti-Spam* Naïve Bayesian

Berikut adalah contoh perhitungan Filter Anti *Spam* menggunakan Naïve Bayesian [3]. Untuk setiap pesan (email) yang direpresentasikan oleh vektor x = (x_1, x_2, …, x_n), dimana x_1, x_2, …, x_n nilai atribut X_1, X_2, …, X_n. Menggunakan atribut biner (Suhami et.al):

Xi =1 jika beberapa karakteristik yang terwakili *Xi* berada dalam pesan; *Xi=0* lainnya.

Dalam percobaan, atribut berkorespondensi dengan kata-kata, yaitu setiap atribut akan menunjukkan jika kata tertentu (misal kata "MORTGAGE") muncul. Untuk memilih dari sekian banyak atribut yang mungkin, kita mengikuti Suhami et.al [3] menghitung *mutual information (MI)* untuk setiap kandidat atribut *X* dengan mendenotasikan kategori variabel *C*:

$$MI(X,C) = \sum_{x \in \{0,1\},\, c \in \{spam,\, legitimate\}} P(X=x, C=c) \cdot \log \frac{P(X=x, C=c)}{P(X=x)\cdot P(C=c)}$$

Atribut dengan nilai *MI* terbesar yang dipilih. Probabilitas ditaksir sebagai rasio frekuensi dari *corpus* yang telah dilatih (lihat Mitchell, 1996, untuk penaksir yang lebih lanjut).

Dari teorema Bayes dan teorema probabilitas total, diberikan vector x = (x_1, x_2, …, x_n), dari dokumen *d*, probabilitas *d* masuk dalam kategori *c* adalah:

$$P(C=c \mid \vec{X} = \vec{x}) = \frac{P(C=c)\cdot P(\vec{X} = \vec{x} \mid C=c)}{\sum_{k \in \{spam,\, legitimate\}} P(C=k)\cdot P(\vec{X} = \vec{x} \mid C=k)}$$

Probabilitas *P(Xi|C)* tidak mungkin, secara praktek, menaksir secara langsung (karena nilai yang mungkin dari $\vec{X}$ terlalu banyak, yang menyebabkan masalah sisa data). Pengklasifikasi Naïve Bayesian membuat pengasumsian yang lebih sederhana sehingga X_1, X_2, …, X_n bebas bersyarat terhadap kategori/kelas *c*. Maka:

$$P(C=c \mid \vec{X} = \vec{x}) = \frac{P(C=c)\cdot \prod_{i=1}^{n} P(X_i = x_i \mid C=c)}{\sum_{k \in \{spam,\, legitimate\}} P(C=k)\cdot \prod_{i=1}^{n} P(X_i = x_i \mid C=k)}$$

dimana *P(Xi|C)* dan *P(C)* dengan mudah dapat ditaksir sebagai frekuensi relatif dari *corpus* yang telah dilatih. Beberapa studi telah menemukan bahwa pengklasifikasi Naïve Bayesian efektif (Langley et al.,1992; Domingos & Pazzani, 1996), walaupun dalam kenyataan bahwa asumsi keterbebasannya biasanya terlalu disederhanakan.

Kesalahan dalam menyaring ham (mengkategorikannya sebagai *spam*) lebih aman dibandingkan meloloskannya (mengkategorikannya sebagai ham) melewati filter. Misal *L -> S* dan *S -> L* mendenotasikan dua tipe kesalahan (error). Asumsikan bahwa *L -> S* adalah λ kali lebih membutuhkan biaya (mahal) dibanding *S -> L* . Suatu pesan dikategorikan sebagai *spam* jika:

$$\frac{P(C=spam \mid \vec{X} = \vec{x})}{P(C=legitimate \mid \vec{X} = \vec{x})} > \lambda$$

Untuk memperluas agar asumsi independent terpenuhi probabilitas penaksiran akurat, pengklasifikasi yang mengadopsi criteria tersebut mencapai hasil yang optimal [4].

 ISBN: 979-756-061-6

Dalam kasus ini,

$$P(C = spam \mid \vec{X} = \vec{x}) = 1 - P(C = legitimate \mid \vec{X} = \vec{x})$$

yang akan membawa pada reformulasi alternative dari kriteria:

$$P(C = spam \mid \vec{X} = \vec{x}) > t \text{ , with } t = \frac{\lambda}{1+\lambda}, \ \lambda = \frac{t}{1-t}$$

Sahami dkk menentukan nilai *t* pada 0.999 (χ =999), yang berarti memblok legitimate pesan sama buruknya dengan membiarkan 999 pesan *spam* melewati filter. Nilai χ yang sedemikian besar tersebut cukup beralasan apabila pesan yang diblok diacuhkan tanpa pemrosesan lebih lanjut, sebagaimana kebanyakan pengguna mempertimbangkan bahwa kehilangan pesan legitimate tidak dapat diterima. Namun memblok pesan *legitimate* lebih "aman" daripada membiarkan pesan *spam* melewati filter.

3.3 Tingkatan Filtering dalam Bayes

Penerapan teori bayes dalam penyaringan email, apakah termasuk dalam kategori *spam* atau bukan, sangatlah akurat, dikarenakan karakteristik dari *spam* tersebut yang akan terulang pada setiap *client*. Karakteristik pengulangan tersebut yang menjadi point dari penggunaan teori Bayes untuk filtering *spam*.

banyak digunakan dalam banyak aplikasi filtering/pendeteksian atau pengklasifikasian *spam* , karena filter Bayesian mempunyai tingkatan tingkatan filtering yang sangat *intim* pada objeknya, yakni pada pasangan *text* corpi, pada objek *spam* dan pada objek *ham.* Filtering yang intim ini ditujukan agar filter Bayesian terbiasa mengenali objeknya terlebih dahulu sehingga bisa dengan langsung mendefinisikan mana *spam* atau bukan *spam*.

Sebagai ilustrasi, jika suatu pesan dipecah menjadi elemen-elemen dengan karakteristik khusus (teks, *tag* HTML, URL, dll), dan elemen-elemen tersebut terjadi berulang-ulang dalam sebuah pesan, maka patut dicurigai bahwa pesan tersebut adalah *spam*.

Secara umum filter Bayesian mengenali pesan (email) berdasarkan pada karakteristik sebagai berikut[2]:

- Kata-kata pada badan suatu pesan, tentu juga pada
- Header (pengirim dan path pesan, dan aspek lainnya seperti
- Kode HTML (misal warna-warna yang digunakan, sebagai contoh:warna merah biasanya sering digunakan untuk subjek pada pesan yang tergolong *spam*)
- Pasangan kata, frase, dan
- Meta Information

4. MEMBANGUN FILTER BAYESIAN

Banyak cara untuk membangun filter Bayesian, tulisan ini mengambil secara umum tahapan-tahapan yang dilakukan dalam pembangunan filter Bayesian (diambil dari banyak tulisan/paper):

1. Pembangunan Database *Spam*
2. Pelatihan filter Bayesian
3. Pemfilteran

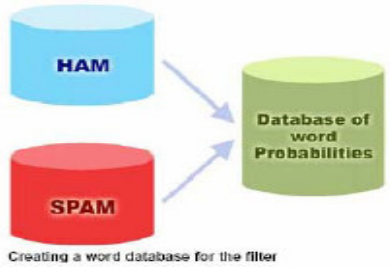
4.1 Pembangunan Database *Spam*

Untuk dapat mengenali karakteristik suatu *spam* diperlukan suatu database kata yang memuat sampel dari kata-kata yang sering dijumpai pada *spam* atau ham. Hal tersebut bertujuan agar filter lebih akurat dalam menjaring *spam*, meminimkan kesalahan dalam pemblokiran ham, hal tersebut seiring dengan banyaknya dan kreatifnya para *spammer* (individu atau kelompok yang mengirim *spam*) dalam mengkotak-katik dan memanipulasi "kata".

4.1.2 Pembuatan database probabilitas kata *(word probabilities database)*

Database probabilitas kata yang berisikan probabilitas kata atau *token* (yang termasuk *spam*) misalnya nama domain,,tanda '$', IP addresses, warna merah, dan lain-lain . Isi mesin database tersebut dikumpulkan dari sampel email *spam* dan email yang valid (*valid email),* selanjutnya disebut *ham.*

Berikut model database bayes yang terdiri dari database *ham, spam,* dan *probabilitas kata*, pada gambar 1.



Gambar 1. Pembuatan DataBase Bayes

Setiap kata atau token yang ada diberi nilai probabilitas; probabilitas tersebut berdasarkan perhitungan, seberapa sering suatu kata muncul dalam *spam* dan berbanding terbalik dengan *ham.* Sebagai contoh perhitungan probabilitas kata; jika kata "MORTGAGE" muncul sebanyak 400 kali dari 3000 email *spam* dan 5 kali dari 300 email sah, sebagai contoh, maka probabilitas email *spam* ini sebesar 0,8889 (didapat dari, [400/3000] dibagi [5/300 + 400/3000]).

4.1.3 Pembuatan database ham

Pembuatan database *ham* juga tidak kalah pentingnya terutama bagi institusi atau perusahaan yang salah satu sarana komunikasinya melalui

ISBN: 979-756-061-6

Internet. Sebagai contoh, suatu perusahaan yang bergerak di bidang keuangan mempunyai tipikal penggunaan kata "mortgage" berkali-kali dan apabila menggunakan anti-*spam* biasa akan berdampak kesalahan positif bagi perusahaan. Namun apabila menggunakan filter Bayesian, kata "MORTGAGE" tersebut menjadi bahan pelatihan , apakah termasuk dalam kategori/kelas *spam* atau ham.

4.1.4 Pembuatan database *spam*

Hal yang penting yaitu pembuatan database *spam*, sebagai bahan pelatihan bagi filter Bayesian dalam mengidentifikasi suatu pesan termasuk dalam kategori *spam*. Kapasitas database *spam* harus memuat sampel *spam* dalam jumlah besar dan terus di *up-date* dengan menggunakan software anti-*spam*. Sehingga filter Bayesian dapat mengidentifikasi dengan lebih cepat dan meningkatkan tingkat keakuratannya serta dapat mengatasi trik-trik terbaru dari *spam*.

4.2 Pelatihan Filter Bayesian

Setelah pembuatan database selesai, tahap selanjutnya adalah pelatihan filter Bayesian agar terbiasa dan *up-to-date* dalam mengidentifikasi atau mendeteksi *spam* atau non-*spam*. Beberapa metode dapat digunakan dalam pelatihan filter Bayesian, tiga diantaranya[1]:

- ***TEFT – Train Everything*** – untuk setiap anggota dari himpunan teks, klasifikasikan teks, rekam/record outputnya (*benar* atau *tidak benar*), dan latih teks tersebut ke dalam database dalam kategori *benar*.
- ***TOE – Train Only Error*** – untuk setiap anggota dari himpunan teks, klasifikasikan teks, rekam outputnya (*benar* atau *tidak benar*), dan jika teks tersebut terklasifikasikan dengan *tidak benar*, maka latih teks tersebut ke dalam database dalam kategori *benar*.
- ***TUNE – Train Until No Errors*** – untuk setiap 500-pesan pertama, klasifikasi ulang dan latih pesan-pesan tersebut jika *tidak benar*. Setelah pengujian pelatihan yang intensif ini dan merekam 500 teks tersebut, maka latih kembali filter Bayesian jika terjadi error sampai tidak adanya error.

Berikut hasil test yang dilakukan pada tabel 1[1] untuk ketiga metode pelatihan di atas, dengan spesifikasi Transmeta 666 dengan RedHat Linux 7.3 dan memori 128 megabytes dengan ukuran dibatasi sampai 1.000.000 slot, dan slot terus menurun apabila tidak ada slots yang tidak digunakan:

Tabel 1. Perbandingan Metode Pelatihan Filter Bayesian

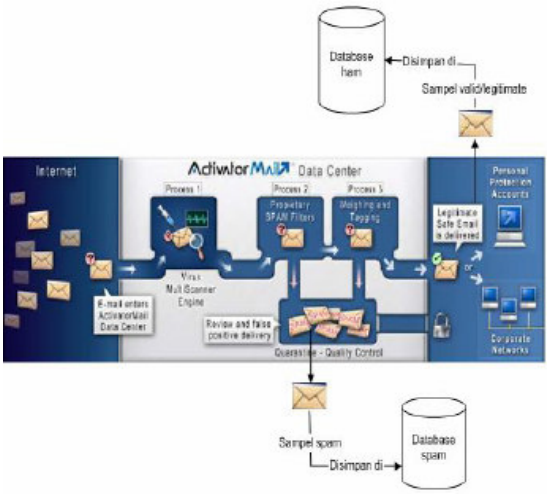
Training Method	Errors per 5000	Approximate time for all 10 shuffles
TEFT	149	6 hrs
TOE	69	3 hrs
TUNE	54	20 hrs

4.3 Pemfilteran

Apabila database ham dan database *spam* telah dibuat, probabilitas kata dapat dihitung dan tersimpan dalam database probabilitas kata, maka filter sudah dapat digunakan.

Ketika suatu pesan (email) dating, maka akan dipecah menjadi kata atau kata yang relevan dengannya. Lebih detilnya, suatu email yang masuk akan diperiksa berdasarkan kata-kata yang termasuk dalam karakteristik tipikal yang telah ditentukan, yaitu badan pesan, headernya, kode HTML, kalimat, frase, *meta information.* Dari kata-kata tersebut kemudian dihitung probabilitas suatu email tergolong *spam* atau non-*spam*. Jika probabilitas melebihi batas toleransi yang ditetapkan sebagai *spam*, misal 0.9, maka email tersebut masuk dalam kategori *spam* dan tidak dapat masuk ke dalam *client inbox*, jika tidak email tersebut masuk dalam kategori *ham* dan masuk ke *client inbox*.

Secara garis besar proses pemfilteran email dari Internet hingga ke dalam *client inbox* digambarkan dalam gambar 2[6].



Gambar 2. Prosedur Pemfilteran Bayesian

5. ANALISIS & KESIMPULAN

Pembentukan database *spam* dan *non-spam* menjadi suatu keharusan agar filter Bayesian selalu waspada dan *up-to-date*dalam mengenali suatu *spam*. Di satu sisi kekuatan anti-*spam* filter Bayesian nampaknya cukup akurat dan signifikan dalam mendeteksi atau "membunuh" *spam,* namun kelebihan tersebut di sisi lain menjadi kelemahan

ISBN: 979-756-061-6

manakala tingkat error atau (false positive) meningkat dalam mengklasifikasikan suatu email *spam* atau non-*spam* , manakala kita menetapkan standar yang terlalu tinggi (mendekati 100%) dalam menggolongkan nilai dari kata-kata dalam email mengandung *spam* atau non-*spam*. Sehingga kita memblokir atau menghapus suatu email yang ternyata bukan tergolong *spam*.

Pendeteksian suatu email masuk dalam kategori *spam* atau *non-spam* membutuhkan suatu pelatihan bagi filter Bayesian agar terbiasa dengan kata-kata yang terkandung dalam email (beberapa probabilitas bahwa kata yang terulang dari populasi di database) semakin baik proses pelatihan maka akan semakin pintar pula fileter Bayesian dalam menyaring suatu spam.

Yang menjadi kunci di sini yaitu,memblok suatu email sebagai spam lebih aman disbanding memloloskannya sebagai ham, dalam rangka menjaga keamanan jaringan komputer pengguna atau institusi.

DAFTAR PUSTAKA

[1] Yerazunis, William S., PhD., "*The Spam-Filtering Accuracy Plateau at 99.9% Accuracy and How to Get Past It"*.

[2] Wikipedia, *"Bayesian filtering"*the free encyclopedia. http://www.wikipedia.org (terakses 03 Februari 2005)

[3] Ion Androutsopoulos, John Koutsias, Konstantinos V. Chandrinos, George Paliouras and Constantine D. Spyropoulos, "*An Evaluation of Naive Bayesian Anti-Spam Filtering"*. Software and Knowledge Engineering Laboratory National Centre for Scientific Research "Demokritos" 153 10 Ag. Paraskevi, Athens, Greece.

[4] Duda, R.O., and Hart, P.E. Bayes Decision Theory. 1973, "*Pattern Classification and Scene Analysis",* Chapter 2, pp. 10–43. John Wiley,.

[5] http://www.gfi.com, *Why Bayesian filtering is the most effective anti-spam technology* (terakses 24 November 2004)

[6] "*How much can your enterprise save using the best antispam and antivirus email security services on the internet*?", http://www.activatormail.com, (terakses 24 November 2004)

[7] Graham, Paul., *BETTER BAYESIAN FILTERING,* http://www.paulgraham.com /better.html, January 2003 (terakses 25 November 2004)