

PENGLASIFIKASIAN E-MAIL UNTUK MENDETEKSI SPAM MENGGUNAKAN ALGORITMA NAIVE BAYESIAN

Ferdi / 0322043

Jurusan Teknik Elektro, Fakultas Teknik, Universitas Kristen Maranatha

Jl. Prof. Drg. Suria Sumantri 65, Bandung 40164, Indonesia

Email : swix4ever@yahoo.com

ABSTRAK

Internet telah menjadi salah satu hal yang penting dalam perkembangan sarana komunikasi. Salah satu fasilitas yang terdapat pada internet adalah surat elektronik atau yang lebih dikenal sebagai *e-mail*. Fasilitas *e-mail* yang mudah digunakan dan murah mengakibatkan banyaknya *e-mail* yang berisi iklan masuk ke dalam inbox pengguna *e-mail*. *E-mail* iklan inilah yang disebut sebagai *spam mail*. Untuk mencegah hal ini, dibuatlah *software* yang berguna sebagai *spam filter* untuk menyaring *e-mail* yang masuk ke dalam *inbox* pengguna fasilitas *e-mail*.

Pemrograman *spam filter* pada tugas akhir ini menggunakan algoritma yang dinamakan *Naive Bayesian*. Metode ini dipilih selain karena merupakan pengembangan terbaru dari pemrograman *spam filter*, juga karena algoritma ini memiliki tingkat keakuratan yang lebih tinggi dibandingkan dengan algoritma sebelumnya (contoh: *NN-Classifier*).

Dengan menggunakan *Visual Basic 6.0*, program *spam filter* dengan algoritma *Naive Bayesian* telah berhasil dibuat. Dengan program ini dilakukan beberapa percobaan yang berhasil membuktikan bahwa algoritma ini mampu mengidentifikasi *spam*, dengan beberapa syarat dan kondisi, secara lebih akurat.

Kata kunci : spam, spam filter, naive bayesian.

E-MAIL CLASSIFICATION FOR SPAM DETECTION USING NAIVE BAYESIAN ALGORITHM

Ferdi / 0322043

Electrical Engineering, Engineering Faculty, Maranatha Christian University

Jl. Prof. Drg. Suria Sumantri 65, Bandung 40164, Indonesia

Email : swix4ever@yahoo.com

ABSTRACT

Internet had become one of the most important thing in growth of communication. One of the facility that internet provide is electronic mail or commonly known as e-mail. The facility of e-mail that easy to use and cheap also makes lot of e-mail that contain advertisement get into the inbox of e-mail user. This advertise mail called spam mail. Avoiding this spam mail, some software are made to be the spam filter that can filter the mail that get into the inbox of e-mail user facility.

To create a spam filter program in this final task, Algorithm that called Naive Bayesian is used. This method choosed, instead of because it is a new method to create a spam filter, it also because that this algorithm have a very high level of accuration to identify spam mail than the previous algorithm (e.g: NN-Classifier).

Using Visual Basic 6.0, a spam filter program with Naive Bayesian algorithm has been made. With this program, some experiment done and successfully proved that this algorithm can identify spam mail, with some term and condition, more accurately.

Keyword : spam, spam filter, naive bayesian.

KATA PENGANTAR

Pertama-tama penulis memanjatkan puji dan syukur kepada Tuhan Yang Maha Esa atas berkat, lindungan dan rahmat-Nya penulis masih diberikan kesempatan untuk dapat menyelesaikan Tugas Akhir ini dengan baik dan tepat pada waktunya.

Tugas Akhir yang berjudul “Pengklasifikasian E-Mail untuk Mendeteksi Spam Menggunakan Algoritma Naive Bayesian” ini dibuat untuk memenuhi persyaratan program studi Strata Satu Jurusan Teknik Elektro Fakultas Teknik Universitas Kristen Maranatha Bandung.

Dalam menyelesaikan Tugas Akhir ini penulis mengalami banyak hambatan karena terbatasnya referensi yang penulis pergunakan, juga karena keterbatasan pengetahuan dan pengalaman penulis. Penulis menyadari bahwa tanpa anugerah Tuhan, serta bantuan dari berbagai pihak, Tugas Akhir ini tidak dapat terselesaikan dengan sempurna. Oleh karena itu penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian Tugas Akhir ini, yaitu :

1. Bpk. Ir. Aan Darmawan, MT., selaku Ketua Jurusan Teknik Elektro Universitas Kristen Maranatha.
2. Ibu Ir. Anita Soepariono, Msc., selaku Koordinator Tugas Akhir Jurusan Teknik Elektro Universitas Kristen Maranatha.
3. Bapak Marvin Chandra Wijaya, ST, MM, MT., selaku Pembimbing Tugas Akhir yang telah turut menyumbangkan pengetahuan dan membimbing penulis hingga Tugas Akhir ini dapat terselesaikan.
4. Mr. John Graham-Cumming yang telah memberikan pencerahan dan kontribusi dalam memilih metoda terbaik dalam pengembangan Tugas Akhir ini.
5. Bpk Agus Sulaiman, ST, MT, yang juga turut menyumbangkan pengetahuannya hingga Tugas Akhir ini dapat terselesaikan dengan baik.

6. Orang tua, saudara, dan teman-teman yang telah memberikan dukungan baik secara moril maupun materiil.

Dengan menyadari bahwa tidak ada segala sesuatu yang sempurna, penulis sangat mengharapkan adanya kritik dan saran yang membangun agar segala kekurangan pada laporan Tugas Akhir ini dapat diperbaiki.

Akhir kata, penulis mengucapkan terima kasih yang sebesar-besarnya kepada seluruh pihak. Penulis berharap laporan ini dapat bermanfaat bagi semua pihak yang membutuhkan.

Bandung, Agustus 2007

Penulis

DAFTAR ISI

Lembar Pengesahan	
Pernyataan Orisinalitas Laporan	
Lembar Pernyataan Persetujuan Publikasi	
Abstrak	i
Abstract	ii
Kata Pengantar	iii
Daftar Isi	v
Daftar Gambar	viii
Daftar Tabel	ix
Daftar Lampiran	x
Bab I Pendahuluan	
I.1 Latar Belakang Masalah	1
I.2 Identifikasi Masalah	1
I.3 Tujuan	1
I.4 Pembatasan Masalah	1
I.5 Sistematika Penulisan	2
Bab II Dasar Teori	
II.1 Pengenalan Visual Basic	3
II.1.1 IDE Visual Basic	3
II.1.2 Jendela IDE	3
II.1.3 ToolBox	5
II.1.4 Mengatur Lingkungan Kerja Visual Basic	6
II.2 Struktur Kendali dan Perulangan	7
II.2.1 If...Then	7
II.2.2 If...Then...Else	7
II.2.3 Select Case	8
II.2.4 Do...Loop	9
II.2.5 For...Next	10
II.2.6 Struktur Kendali Nested	10

II.3	Variabel dan Operator pada Visual Basic	10
II.3.1	Aturan Penamaan Variabel	11
II.3.2	Tipe Variabel	11
II.3.3	Operator Pada Visual Basic dan Urutan Operasinya	12
II.4	Algoritma Naive Bayesian	13
II.4.1	Rumus Bayes	13
II.4.2	Pengaplikasian Naive Bayesian	13
Bab III Perancangan Program		
III.1	Tampilan Program	15
III.1.1	Form Menu Utama	15
III.1.2	Form Menu Input ke Database (Learning)	17
III.1.3	Form Check Database	19
III.1.4	Form Klasifikasi	21
III.2	Flowchart Program	22
III.2.1	Flowchart Menu Utama	23
III.2.2	Flowchart Input Database (Learning)	23
III.2.3	Flowchart Input ke Database Spam	24
III.2.4	Flowchart Input ke Database Bukan Spam	25
III.2.5	Flowchart Klasifikasi File E-mail	26
III.3	Listing Program	29
III.3.1	Program Input ke Database Spam	29
III.3.2	Program Input ke Database Bukan Spam	30
III.3.3	Program Koneksi Database	31
III.3.4	Program Klasifikasi	31
III.4	Pemrograman Database	35
Bab IV Data Pengamatan dan Analisa Data		
IV.1	Tampilan Program	37
IV.1.1	Input ke Database Spam	37
IV.1.2	Input ke Database Bukan Spam	39
IV.1.3	Tampilan Klasifikasi	41
IV.1.3.1	Klasifikasi Spam	41
IV.1.3.2	Klasifikasi Bukan Spam	42

IV.1.3.3 Klasifikasi Bukan Keduanya	43
IV.2 Penambahan Bukan Spam	44
IV.2.1 Menggunakan 5 File Spam dan 5 File Bukan Spam	44
IV.2.2 Menggunakan 5 File Spam dan 8 File Bukan Spam	45
IV.2.3 Menggunakan 5 File Spam dan 10 File Bukan Spam	46
IV.3 Penambahan Spam	47
IV.3.1 Menggunakan 10 File Spam dan 5 File Bukan Spam	47
IV.3.2 Menggunakan 15 File Spam dan 5 File Bukan Spam	49
IV.3.3 Menggunakan 19 File Spam dan 5 File Bukan Spam	51
IV.3.4 Menggunakan 21 File Spam dan 5 File Bukan Spam	52
IV.4 Percobaan Pembuktian Kesebandingan	54
IV.4.1 Dengan 1 File Spam dan 1 File Bukan Spam	54
IV.4.2 Dengan 2 File Spam dan 1 File Bukan Spam	55
IV.4.3 Dengan 1 File Spam dan 2 File Bukan Spam	55
IV.5 Analisa Data Pengamatan	56
Bab V Kesimpulan dan Saran	
V.1 Kesimpulan	57
V.2 Saran	57
Daftar Pustaka	58
Lampiran	xi

DAFTAR GAMBAR

Gambar II.1	IDE <i>Visual Basic</i> dengan jendela-jendela yang terbuka	4
Gambar III.1	Tampilan <i>Menu Utama</i>	15
Gambar III.2	Tampilan <i>Menu Input ke Database</i>	17
Gambar III.3	III.3 Tampilan <i>Menu Check Database</i>	19
Gambar III.4	III.4 Tampilan <i>Menu Klasifikasi</i>	21
Gambar III.5	<i>Flowchart Menu Utama</i>	23
Gambar III.6	<i>Flowchart Proses Database Input</i>	23
Gambar III.7	<i>Flowchart Input ke Database Spam</i>	24
Gambar III.8	<i>Flowchart Input ke Database Bukan Spam</i>	25
Gambar III.9	<i>Flowchart Pengklasifikasian File E-mail (bag.1)</i>	26
Gambar III.10	<i>Flowchart Pengklasifikasian File E-mail (bag.2)</i>	27
Gambar III.11	<i>Flowchart Pengklasifikasian File E-mail (bag.3)</i>	28
Gambar IV.1	<i>Input ke Database Spam</i>	37
Gambar IV.2	Hasil <i>Input</i> pada <i>Database Spam</i>	38
Gambar IV.3	<i>Data Spam</i> yang dapat dilihat melalui <i>Check Database</i>	38
Gambar IV.4	<i>Input ke Database Bukan Spam</i>	39
Gambar IV.5	Hasil <i>Input</i> pada <i>Database Bukan Spam</i>	40
Gambar IV.6	<i>Data Bukan Spam</i> yang dapat dilihat melalui <i>Check Database</i>	40
Gambar IV.7	Hasil <i>Klasifikasi Spam</i>	41
Gambar IV.8	Hasil <i>Klasifikasi Bukan Spam</i>	42
Gambar IV.9	Hasil <i>Klasifikasi Bukan Kedua Kategori</i>	43

DAFTAR TABEL

Tabel II.1	Tabel Tipe Variabel-Ukuran-dan Kisarannya	12
Tabel II.2	Tabel Operator Visual Basic	12
Tabel III.1	Daftar Komponen <i>Form Menu</i> Utama	16
Tabel III.2	Daftar Komponen <i>Form Menu Input</i> ke <i>Database</i>	18
Tabel III.3	Daftar Komponen <i>Form Menu Check Database</i>	19
Tabel III.4	Daftar Komponen <i>Form Menu</i> Klasifikasi	22
Tabel III.5	Pemrograman Tabel dalam <i>Database</i>	35
Tabel IV.1	Percobaan dengan menggunakan 5 <i>File Spam</i> dan 5 <i>File Bukan Spam</i>	44
Tabel IV.2	Percobaan dengan menggunakan 5 <i>File Spam</i> dan 8 <i>File Bukan Spam</i>	45
Tabel IV.3	Percobaan dengan menggunakan 5 <i>File Spam</i> dan 10 <i>File Bukan Spam</i>	46
Tabel IV.4	Percobaan dengan menggunakan 10 <i>File Spam</i> dan 10 <i>File Bukan Spam</i>	48
Tabel IV.5	Percobaan dengan menggunakan 15 <i>File Spam</i> dan 10 <i>File Bukan Spam</i>	50
Tabel IV.6	Percobaan dengan menggunakan 19 <i>File Spam</i> dan 10 <i>File Bukan Spam</i>	51
Tabel IV.7	Percobaan dengan menggunakan 21 <i>File Spam</i> dan 10 <i>File Bukan Spam</i>	53
Tabel IV.8	Kesebandingan 1	55
Tabel IV.9	Kesebandingan 2	55
Tabel IV.10	Kesebandingan 3	55

DAFTAR LAMPIRAN

A.1	Form1.....	xi
A.2	Form2.....	xii
A.3	Form3.....	xv
A.4	Form4.....	xix
B.1	Bukan Spam 0 (Bukan 0).....	xx
B.2	Bukan Spam 1 (Bukan 1)	xx
B.3	Bukan Spam 2 (Bukan 2)	xxi
B.4	Bukan Spam 3 (Bukan 3)	xxi
B.5	Bukan Spam 4 (Bukan 4)	xxiii
B.6	Bukan Spam 5 (Bukan 5)	xxiii
B.7	Bukan Spam 6 (Bukan 6)	xxiv
B.8	Bukan Spam 7 (Bukan 7)	xxv
B.9	Bukan Spam 8 (Bukan 8)	xxvi
B.10	Bukan Spam 9 (Bukan 9)	xxviii
C.1	Spam 0.....	xxix
C.2	Spam 1.....	xxix
C.3	Spam 2.....	xxix
C.4	Spam 3.....	xxx
C.5	Spam 4.....	xxx
C.6	Spam 5.....	xxxi
C.7	Spam 6.....	xxxi
C.8	Spam 7.....	xxxi
C.9	Spam 8.....	xxxii
C.10	Spam 9.....	xxxii