



**ANALISA PENERAPAN DATA MINING DALAM EMAIL
FILTERING MENGGUNAKAN METODE NAÏVE BAYESIAN**

TUGAS AKHIR

YUNI RAMADHINI

41506110034

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2010**



**ANALISA PENERAPAN DATA MINING DALAM EMAIL
FILTERING MENGGUNAKAN METODE NAÏVE BAYESIAN**

Laporan Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

YUNI RAMADHINI

41506110034

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2010**

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NIM : 41506110034

Nama : Yuni Ramadhini

Judul Skripsi : ANALISA PENERAPAN DATA MINING DALAM EMAIL
FILTERING MENGGUNAKAN METODE NAÏVE BAYESIAN

Menyatakan bahwa skripsi tersebut diatas adalah hasil karya saya sendiri dan bukan plagiat. Apabila ternyata ditemukan di dalam laporan skripsi saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, Februari 2010

Yuni Ramadhini

LEMBAR PERSETUJUAN

Nim : 41506110034
Nama : YUNI RAMADHINI
Judul Skripsi : ANALISA PENERAPAN DATA MINING DALAM
EMAIL FILTERING DENGAN MENGGUNAKAN
METODE NAIVE BAYESIAN

SKRIPSI INI TELAH DIPERIKSA DAN DISETUJUI
JAKARTA, FEBRUARI 2010

E. Didik Madyatmadja, ST., M.Kom
Pembimbing

Abdusy Syarif, ST., MT.
MTi.
Ketua Program Studi
Teknik Informatika

Devi Fitriannah, S.Kom.
Koordinator Tugas Akhir
Teknik Informatika

KATA PENGANTAR

Bismillaahirrahmaanirrahiim,

Alhamdulillah puji dan syukur penulis ucapkan kepada Allah SWT atas segala nikmat dan karunia-Nya yang telah diberikan, sehingga penulis bisa menyelesaikan tugas akhir ini. Penulis menyadari bahwa dalam penelitian dan penulisan tugas ini tidak terlepas dari bantuan, dukungan, dorongan, kerjasama maupun bimbingan dari berbagai pihak. Untuk itu, penulis mengucapkan terima kasih yang sebesar – besarnya kepada :

1. Bapak Abdusy Syarif, ST., MT selaku Ketua Program Studi Teknik Informatika Universitas Mercubuana.
2. Ibu Devi Fitriana, S.Kom, MTi dan bapak Anis Cherid, S.Kom., M.Ti selaku Koordinator Tugas Akhir Teknik Informatika Universitas Mercubuana.
3. Bapak E. Didik Madyatmadja, ST., M.Kom selaku dosen pembimbing tugas akhir atas bimbingan dan kesabaran yang telah diberikan kepada penulis dari awal sampai selesainya tugas akhir ini.
4. Segenap staf pengajar di fakultas Ilmu Komputer Universitas Mercubuana yang telah mengajar, membimbing dan memberikan pemahaman-pemahaman penulis tentang ilmu komputer dan informatika.

5. Tommi Mauliddin dan Almer Rafif Azaria, suami dan anak tercinta yang terus menemani dan memberikan semangat tiada henti kepada penulis.
 6. Keluarga Purwoto yang telah memberikan semangat dan doa kepada penulis dalam mengerjakan tugas akhir.
 7. Burhanuddin, Agung, Dany, Sri Utami, Diah dan teman – teman yang selalu membantu diskusi dan memberi dukungan.
 8. Semua pihak dan teman-teman yang telah membantu penulis dalam memberikan doa dan semangat sehingga selesainya pengerjaan tugas akhir ini.
- Penulis berharap laporan tugas akhir ini dapat bermanfaat bagi yang membutuhkannya.

Jakarta, Februari 2010

Penulis

DAFTAR ISI

LEMBAR PERNYATAAN	ii
LEMBAR PERSETUJUAN	iii
KATA PENGANTAR.....	iv
<i>ABSTRACTION</i>	vi
ABSTRAKSI	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Ruang Lingkup	4
1.4 Tujuan Pembahasan	4
1.5 Metode Penyelesaian Masalah.....	4
1.6 Sistematika Penulisan	5
BAB II LANDASAN TEORI	6
2.1 Algoritma Bayesian	6
2.1.1 Perhitungan Probabilitas dengan Algoritma Bayesian.....	8
2.1.2 Metode <i>Tokenizing</i>	11
2.1.3 Metode Scoring.....	11

2.1.4 Metode <i>Combining</i>	11
2.2 <i>Microsoft Office Outlook</i> 2003.....	12
2.3 <i>Microsoft Visual Studio</i> 2008.....	13
2.4 <i>Microsoft .Net Framework</i>	14
BAB III ANALISA DAN PERANCANGAN SISTEM	17
3.1 Analisa Sistem.....	17
3.1.1 Proses <i>Training</i>	18
3.1.2 Proses <i>Tokenizing</i>	19
3.1.3 Proses <i>Scoring</i>	19
3.1.4 Proses <i>Combining</i>	21
3.2 Perancangan Sistem.....	22
3.2.1 <i>Use case Diagram</i>	22
3.2.2 <i>Class Diagram</i>	24
3.2.3 <i>Activity Diagram</i>	27
3.2.4 <i>Sequence Diagram</i>	30
3.2.5 Proses Inti <i>Spam Filter</i>	34
3.2.6 Perancangan <i>User Interface</i>	34
BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM	38
4.1 Implementasi.....	38
4.1.1 Perangkat Keras	38
4.1.2 Perangkat Lunak	39
4.1.3 Implementasi Program Antar Muka	39

4.1.3.1 Antar Muka Menu Utama	39
4.1.3.2 Antar Muka Menu <i>Option</i>	39
4.1.3.3 Antar Muka Menu <i>Analysis</i>	41
4.1.3.4 Antar Muka Menu <i>Tokens</i>	43
4.1.3.5 Antar Muka Menu <i>Rebuild</i>	44
4.1.4 Implementasi Rancangan <i>Class</i>	44
4.2 Pengujian	50
4.2.1 Tujuan Pengujian	50
4.2.2 Rancangan Pengujian	51
4.2.3 Langkah – Langkah Pengujian	51
4.2.4 Hasil Pengujian	51
4.2.5 Analisis Hasil Pengujian.....	56
BAB V PENUTUP	57
5.1 Kesimpulan.....	57
5.2 Saran.....	57
DAFTAR PUSTAKA	59
LAMPIRAN A FLOWCHART NILAI PROBABILITAS TOKEN.....	60
LAMPIRAN B NILAI PROBABILITAS TOKEN	61
LAMPIRAN C LISTING PROGRAM	64

DAFTAR GAMBAR

Gambar 3.1 <i>Flowchart</i> Utama Program <i>Spam Filter</i>	18
Gambar 3.2 <i>Flowchart</i> set to spam mail	20
Gambar 3.3 <i>Flowchart</i> set to ham mail	22
Gambar 3.4 <i>Use case</i> Diagram	23
Gambar 3.5 <i>Class</i> Diagram	25
Gambar 3.6 <i>Activity</i> Diagram <i>Add To Good</i>	27
Gambar 3.7 <i>Activity</i> Diagram <i>Add To Spam</i>	28
Gambar 3.8 <i>Activity</i> Diagram <i>Hunt Spam</i>	29
Gambar 3.9 <i>Sequence</i> Diagram <i>Add To Good</i>	31
Gambar 3.10 <i>Sequence</i> Diagram <i>Add To Spam</i>	32
Gambar 3.11 <i>Sequence</i> Diagram <i>Hunt Spam</i>	33
Gambar 3.12 Rancangan Awal <i>Toolbar</i>	35
Gambar 3.13 Rancangan <i>Form Spam Analysis</i>	35
Gambar 3.14 Rancangan <i>Form</i> untuk <i>option</i>	37
Gambar 3.15 Rancangan tampilan untuk <i>table token</i>	37
Gambar 4.16 <i>Toolbar EmailFilter</i>	39
Gambar 4.17 <i>Form option</i>	40
Gambar 4.18 Kotak Dialog <i>Select Folder</i>	40

Gambar 4.19 <i>Form Analysis Mail Text</i>	41
Gambar 4.20 <i>Form Analysis Token dalam Mail</i>	42
Gambar 4.21 <i>Form Analysis 15 Token</i>	42
Gambar 4.22 <i>Tokens</i>	43
Gambar 4.23 <i>Message Dialog Konfirmasi Rebuild</i>	43
Gambar 4.24 <i>Tampilan hasil program Data 15 Token</i>	54
Gambar 4.25 <i>Tampilan Hasil Spam Level</i>	55
Gambar Lampiran A.26 <i>Flowchart nilai probabilitas token</i>	60

DAFTAR TABEL

Tabel 1.1 Perbandingan performansi metode – metode anti <i>spam filter</i>	2
Tabel 4.2 <i>Method Class ThisAddIn</i>	44
Tabel 4.3 <i>Method Class Corpus</i>	46
Tabel 4.4 <i>Method Class SpamFilter</i>	47
Tabel 4.5 <i>Method Class FormAnalysis</i>	48
Tabel 4.6 <i>Method Class FormTokenTable</i>	49
Tabel 4.7 <i>Method Class FormOption</i>	49
Tabel 4.8 Pengujian perubahan data jumlah <i>token</i>	52
Tabel 4.9 Data 15 <i>Token</i>	53
Tabel Lampiran B.10 Nilai Probabilitas Token.....	61