

**DETEKSI SURAT ELEKTRONIK SPAM
MENGUNAKAN JARINGAN SYARAF TIRUAN**

TESIS



Oleh:

S A K U R

1011600879

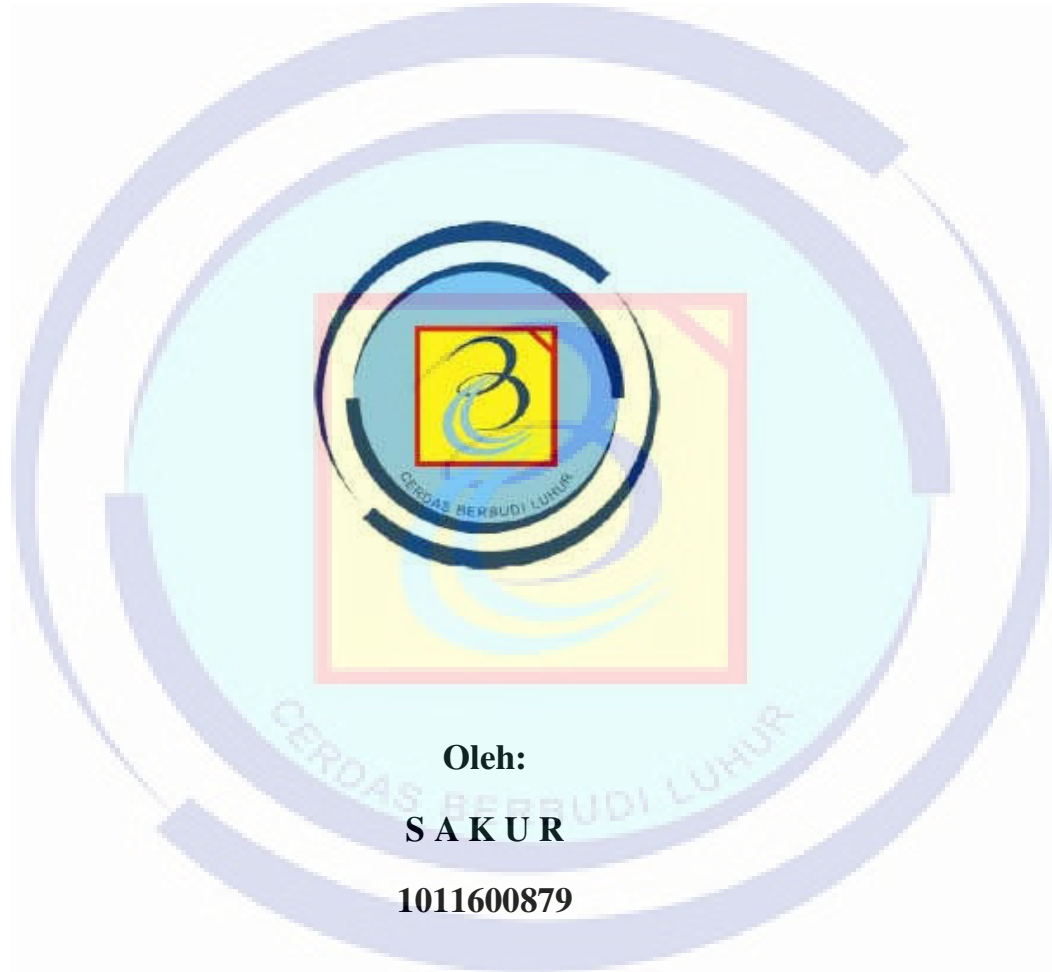
**PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR**

**JAKARTA
2013**

**DETEKSI SURAT ELEKTRONIK SPAM
MENGUNAKAN JARINGAN SYARAF TIRUAN**

TESIS

**Diajukan untuk memenuhi salah satu persyaratan
memperoleh gelar Magister Ilmu Komputer (MKOM)**



Oleh:

S A K U R

1011600879

**PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR**

**JAKARTA
2013**



PROGRAM STUDI ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR

LEMBAR PENGESAHAN

Nama Mahasiswa : S a k u r
NIM : 1011600879
Konsentrasi : Rekayasa Komputasi Terapan
Jenjang Studi : Strata 2
Topik/Judul tesis : Deteksi Surat Elektronik *Spam* Menggunakan Jaringan Syaraf Tiruan

Jakarta, 12 Februari 2013

Tim Penguji

Tanda Tangan

Ketua
Dr. Moedjiono, M.Sc

Anggota
Ir. Sarmidi, M.Kom., M.M

Pembimbing
Dr. Nazori AZ, M.T

Pembimbing
Assoc. Prof. Dr. Anton Satria Prabuwono, M.Sc., M.M.

Ketua Program Studi

Dr. Nazori AZ, M.T



PROGRAM STUDI ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR

LEMBAR PERNYATAAN

Nama Mahasiswa : SAKUR
NIM : 1011600879
Konsentrasi : Rekayasa Komputasi Terapan
Jenjang Studi : Strata 2
Fakultas/Program : Magister Ilmu Komputer

menyatakan bahwa Tesis yang berjudul:

DETEKSI SURAT ELEKTRONIK SPAM
MENGUNAKAN JARINGAN SYARAF
TIRUAN

1. merupakan hasil karya tulis ilmiah sendiri dan bukan merupakan karya yang pernah diajukan untuk memperoleh gelar akademik oleh pihak lain,
2. saya ijin untuk dikelola oleh Universitas Budi Luhur sesuai dengan norma hukum dan etika yang berlaku.

Pernyataan ini saya buat dengan penuh tanggung jawab dan saya bersedia menerima konsekuensi apapun sesuai aturan yang berlaku apabila di kemudian hari pernyataan ini tidak benar.

Jakarta, 12 Februari 2013



(..... SAKUR)

ABSTRAK

Perkembangan penggunaan surat elektronik dalam beberapa tahun terakhir terus meningkat. Sejalan dengan itu peningkatan jumlah surat elektronik *spam* terjadi secara eksponensial. Meningkatnya surat elektronik *spam* ini menyebabkan pemakaian pita lebar meningkat sehingga meningkatkan pengeluaran biaya lebar pita. Tujuan penelitian ini adalah menghasilkan klasifikasi surat elektronik *spam* dan *ham*. Pola karakteristik surat elektronik *spam* ini selalu berubah sehingga diperlukan metode yang dapat digunakan untuk mengenali polanya. Metode yang akan digunakan dalam penelitian ini adalah Jaringan Syaraf Tiruan dengan algoritma propagasi balik. Metode klasifikasi surat elektronik dengan cara mengekstrak kata, kemudian dibobotkan yang menjadi masukan di jaringan syaraf. Dengan menggunakan jaringan syaraf tiruan ini pendeteksian surat elektronik *spam* ini menghasilkan tingkat keakuratan 98%.

Kata kunci: surat elektronik, *spam*, *ham*, propagasi balik, Jaringan Syaraf Tiruan

ABSTRACT

The number of email users is increasing lately. In line with that, a number of spam email shows gradual increase in exponential. As a result, the broadband users also increase the amount of money spent for bandwidth. The aim of this research is to classify spam emails and ham. As the characteristic pattern of spam email is always changing, it is necessary to propose a model that recognizes the pattern. The research method employed in this study is artificial neural network with algorithm back propagation. This classification email method applied by extracting the words and counting them as inputs in neural network. By employing artificial neural network, the detection of spam email results in level accuracy of 98%.

Keyword: Email, spam, ham, back propagation algorithm, artificial Neural Network

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT atas segala rahmat dan perkenan-Nya sehingga proposal tesis ini dapat diselesaikan. Tujuan dari penulisan proposal tesis ini adalah untuk melakukan penelitian sebagai satu syarat untuk menyelesaikan Program Strata Dua (S-2) pada Program Magister Ilmu Komputer Universitas Budi Luhur.

Perjalanan penulis dalam mengikuti program Magister dan menyusun tesis ini tidak terlepas dari dukungan mereka yang terkait dengan penulis langsung maupun tidak langsung. Untuk itu penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Assoc. Prof. Dr. Anton SatriaPrabuwono, M.Sc., M.M., selaku pembimbing yang telah memberikan masukan tentang penyusunan proposal tesis ini.
2. Bapak Dr. Nazori Az. MT, selaku pembimbing kedua sekaligus sebagai Ketua Program Studi Magister Ilmu Komputer yang telah memberikan kemudahan kepada penulis dalam proses pembelajaran di kelas maupun penyusunan proposal ini.
3. Istri dan anakku yang selalu sabar dalam membantu ayahnya baik secara moral maupun materil.
4. Teman-teman program Magister angkatan 2011, terima kasih atas dukungannya.

Penulis Pihak-pihak yang telah membantu dalam penyusunan proposal ini yang belum disebutkan namanya, saya ucapkan terima kasih.

Jakarta Januari 2013

S a k u r



DAFTAR ISI

Halaman

ABSTRAK	i
ABSTRACT	ii
KATA PENGANTAR	iii
DAFTAR ISI	iv
DAFTAR GAMBAR	v
DAFTAR TABEL	vi
DAFTAR LAMPIRAN	vii
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Masalah Penelitian	2
1.2.1 Identifikasi Masalah	2
1.2.2 Batasan Masalah	2
1.2.3 Rumusan Masalah	3
1.3 Tujuan dan Manfaat Penelitian	3
1.3.1 Tujuan Penelitian	3
1.3.2 Manfaat Penelitian	3
1.4 Tata-Urut Penulisan	3
1.5 Daftar Pengertian	4
BAB II LANDASAN PEMIKIRAN	
2.1 Tinjauan Pustaka	6
2.1.1 Pengertian Surat Elektronik	6
2.1.2 Pengertian <i>Spam</i>	7
2.1.3 Perbedaan <i>Spam</i> dan <i>Ham</i>	8
2.1.4 Jaringan Syaraf Tiruan	9
2.1.5 Algoritma Propagasi Balik	11
2.2 Tinjauan Studi	13
2.3 Tinjauan Obyek Penelitian	16
2.3.1 Perangkat Keras	16

2.3.2 Perangkat Lunak	17
2.4 Pola Pikir	18
2.5 Hipotesis	19
BAB III DESAIN PENELITIAN	
3.1 Metode Penelitian.....	20
3.2 Metode Pengumpulan Data.....	20
3.3 Instrumentasi.....	20
3.4 Teknik Analisis Data	21
3.5 Langkah-langkah Penelitian	21
3.6 Jadwal Penelitian.....	24
BAB IV ANALISIS, INTERPETASI DAN IMPLIKASI PENELITIAN	
4.1 Permodelan Proses	25
4.1.1 Pra Proses.....	26
4.1.2 Pascaproses	27
4.2 Pengembangan Aplikasi.....	29
4.2.1 Perancangan data.....	29
4.2.2 Flowchart jaringan syaraf tiruan	33
4.3 Pengujian dan evaluasi aplikasi.....	36
4.3.1 Hasil Percobaan.....	37
4.4 Implikasi Penelitian.....	43
4.4.1 Aspek Sistem.....	43
4.4.2 Aspek Manajerial	44
4.4.3 Aspek Penelitian Lanjut.....	44
4.5 Rencana Implementasi	44
BAB V PENUTUP	46
DAFTAR PUSTAKA	47
LAMPIRAN-LAMPIRAN	

DAFTAR GAMBAR

Gambar	Halaman
II-1	Cara Kerja Surat Elektronik 6
II-2	Neuron pada Saraf Manusia 10
II-2	Algoritma Propagasi Balik 12
II.3	Skema Pola Pikir..... 18
III.1	Langka-Langkah Penelitian..... 21
IV-1	Permodelan proses klasifikasi surat elektronik 25
IV-2	Tampilan Program membuat list kata 29
IV-3	Tampilan program pembuatan Matriks..... 31
VI-4	Matriks 32
IV-5	Flowchart jaringan syaraf tiruan..... 33
IV-6	<i>Neural Network Training</i> 35

DAFTAR TABEL

Tabel	Halaman
1. Daftar Penelitian bidang Surat elektronik <i>Spam</i>	16
2. Jadwal Penelitian.....	24
3. Kata dan jumlah kemunculannya	29
4. Data Spam dengan 5 Neuron	36
5. Data <i>Precision</i> dan <i>Recall</i> dengan 5 Neuron	36
6. Data Spam dengan 10 Neuron.....	37
7. Data <i>Precision</i> dan <i>Recall</i> dengan 10 Neuron.....	37
8. Data Spam dengan 15 Neuron.....	38
9. Data <i>Precision</i> dan <i>Recall</i> dengan 15 Neuron.....	39
10. Data Spam dengan 20 Neuron.....	39
11. Data <i>Precision</i> dan <i>Recall</i> dengan 20 Neuron.....	40
12. Rencana Implementasi.....	46

DAFTAR LAMPIRAN

Lampiran	Halaman
1 Kode Sumber	50
2 Daftar Riwayat Hidup	57



DAFTAR PUSTAKA

- [Arif 2009] Arif, Iwan, **Text Mining**, <http://lecturer.eepisits.edu/~iwanarif/kuliah/dm/6Text%20Mining.pdf>, (Diakses 24 Januari 2013)
- [Almeida 2011] Almeida, Tiago. Almeida, Jurandy.Yamakami, Akebo., **Spam Filtering: How The Dimensionality Reduction Affects The Accuracy Of Naive Bayes Classifiers**, Journal of Internet Services and Applications, Springer, London , February 2011
- [Aery 2005] M. Aery dan S. Chakravarthy., **Emailsift: Email Classification Based On Structure And Content**, In Proc. of 5th IEEE Intl. Conf. on Data Mining, 2005, pp. 1-8.
- [CAN-SPAM 2009] CAN-SPAM Act: A Compliance Guide for Business: 2009:<http://www.business.ftc.gov/documents/bus61-can-spam-act-compliance-guide-business>. (Diakses 13 Desember 2012)
- [Cohen 1996] Cohen, W. W., **Learning Rules That Classify E-Mail**. In AAAI Spring Symposium on Machine Learning in Information Access., 1996
- [Cui 2005] Bin Cui, Anirban Mondel, Jielie Shen, Gao Cong, dan Kian Lee Tan., **On Effective E-mail Classification via Neural Networks**. International Conference on Database and Expert Systems Applications DEXA, 2005
- [CS 2009] Commision Study., **Data Protection: Junk E-Mail Costs Internet Users 10 Billion A Year Worldwide - Commission study**”, <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/01/154>. (diakses 31 Juli 2012)
- [Devan 2000] Devan Wils., **Extreme Programming: A gentle introduction**, <http://www.extremeprogramming.org/> . 2000. (Diakses 01 Januari 2013)
- [Firte 2010] Firte, L., Lemnaru, C., Potolea, R., **Spam detection filter using KNN algorithm and resampling** Intelligent Komputer Communication and Processing (ICCP), IEEE International Conference 2010 , p. 27 - 33
- [Graham 2002] Paul Graham., **Plan For Spam**, <http://www.paulgraham.com/better.html>, 2002 (akses 31 Agustus 2012)
- [Habra 2005] Abdul Habra., **An Introduction Neural Networks**. <http://www.tek271.com/documents/others/into-to-neural-networks>, 2005 (akses 8 Agustus 2012)
- [Hung 2009] Chih-Hung Wu., **Behavior-Based Spam Detection Using A Hybrid Method Of Rule-Based Techniques And Neural Networks Expert**

- Sistems With Applications.** An International Journal archive 2009 p. 4321-4330
- [Inpres 2001] **Kiat Pembakuan Peristilahan Perkomputeran dalam Bahasa Indonesia**, Instruksi Presiden Nomor 2 Tahun 2001 tentang penggunaan komputer dengan aplikasi komputer berbahasa Indonesia, 2001
- [Jong 2004] Jong Jek Siang, M.Sc, Drs., **Jaringan Saraf Tiruan & Pemrogramannya Menggunakan Matlab**. Yogyakarta: Andi. 2004
- [Kusumadewi 2003] Sri Kusumadewi., **Artificial Intelligence (Teknik dan Aplikasinya)**., Yogyakarta: Graha Ilmu, 2003
- [Magdalena 2007] Magdalena, Merry. **How to Read Email Headers**. <http://www.emailaddressmanager.com/tips/email.html>, 2007 (Diakses 25 Januari 2013)
- [Maseleno 2003] Andino Maseleno. **Kamus Istilah Komputer dan Informatika**. <http://ikc.dinus.ac.id/pengantar/andino-kamusti.php>, 2003
- [Mayor 2011] Dedy Mayor Sp., **Analisis dan Implementasi Deteksi Email Spam Menggunakan Karakter N-Grams.**, Skripsi, STT Telkom, 2011
- [Mawarni 2011] Erwin Restika Marwani., **Pendeteksi Konten Email Spam Menggunakan Metode Learning Vector Quantization Berdasarkan Pendekatan Neural Network**, Skripsi, Institut Teknik Surabaya, 2011
- [Møller 1990] Møller, M.F., **Learning by Conjugate Gradients, The 6th International Meeting of Young Komputer Scientists**, Czechoslovakia, in press, 1990
- [Noi 2009] Lucia Di Noi., **Web Spam Detection by Neural Networks for Structures**, Thesis, Universita degli Studi di Siena 2009
- [Prabuwono 2012] Anton Satria Prabuwono., **Neural Network Algorithm**, Diktat Kuliah Kecerdasan Buatan, Universitas Budi Luhur, 2012
- [Puniškis 2006] D. Puniškis, R., Laurutis, R., Dirmeikis., **An Artificial Neural Nets for Spam e-mail Recognition**. Electronics and Electrical Engineering. – Kaunas: Technologija, 2006. p. 73–76.
- [Rafiqul 2005] Md. Rafiqul Islam dan M. U. Chowdhury., **Spam filtering using ML algorithms**, In Proc. of IADIS International Conf. on WWW/Internet, 2005, pp. 419-426.
- [Tran 2010] Tich Phuoc Tran, Thi Thanh Sang Nguyen, Poshiang Tsai, Xiaoying Kong., **BSPNN: Boosted Subspace Probabilistic Neural Network For Email Security.**, Springer Science+Business Media, 2010
- [Tseng 2011] Chi-Yao Tseng; Pin-Chieh Sung; Ming-Syan Chen Cosdes., **A Collaborative Spam Detection Sistem with a Novel E-Mail Abstraction**

Scheme Knowledge and Data Engineering, IEEE Transactions, 2011. p. 669 – 682.

[Tseng 2009] Chi-Yao Tseng, Ming-Syan Chen., **Incremental SVM Model for Spam Detection on Dynamic Email**. International Conference on Social Networks, 2009 p.128 - 135

[Sakkis 2003] Sakkis G, Androutsopoulos I, Paliouras G, Karkaletsis V, Spyropoulos CD, Stamatopoulos P., **A Memory- Based Approach To Anti-Spam Filtering For Mailing Lists**. In: Information retrieval, vol 6. Kluwer Publishing, 2003. p. 49–73

[Wiki 2012] Wikipedia. **Artificial Neural.**, http://en.wikipedia.org/wiki/Artificial_Neural_Network (akses tanggal 31 Juli 2012)

