

**IMPLEMENTASI NAIVE BAYESIAN CLASSIFIER UNTUK
KASUS FILTERING SMS SPAM**

Skripsi



oleh
GILANG JALU SELO WINOYO TUR
22084560

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2013

IMPLEMENTASI NAIVE BAYESIAN CLASSIFIER UNTUK KASUS FILTERING SMS SPAM

Skripsi



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

GILANG JALU SELO WINOYO TUR
22084560

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2013

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

IMPLEMENTASI NAIVE BAYESIAN CLASSIFIER UNTUK KASUS FILTERING SMS SPAM

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 16 Januari 2013



GILANG JALU SELO WINOYO TUR
22084560

HALAMAN PERSETUJUAN

Judul Skripsi : IMPLEMENTASI NAIVE BAYESIAN CLASSIFIER
UNTUK KASUS FILTERING SMS SPAM

Nama Mahasiswa : GILANG JALU SELO WINOYO TUR

N I M : 22084560

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

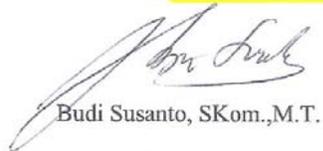
Semester : Gasal

Tahun Akademik : 2012/2013



Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 16 Januari 2013

Dosen Pembimbing I Dosen Pembimbing II


Budi Susanto, S.Kom., M.T.


Rosa Delima, S.Kom., M.Kom.

HALAMAN PENGESAHAN

IMPLEMENTASI NAIVE BAYESIAN CLASSIFIER UNTUK KASUS FILTERING SMS SPAM

Oleh: GILANG JALU SELO WINOYO TUR / 22084560

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 10 Januari 2013

Yogyakarta, 16 Januari 2013
Mengesahkan,

Dewan Penguji:

1. Budi Susanto, SKom., M.T.
2. Rosa Delima, S.Kom., M.Kom.
3. Junius Karel, M.T.
4. Yuan Lukito, S.Kom

DUTA WACANA

Dekan

Ketua Program Studi



(Drs. Wimmie Handwidjojo, MIT.)

(Nugroho Agus Haryono, M.Si)

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepada Tuhan Yang Mahaesa atas berkat, rahmat, dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Implementasi Naïve Bayesian Classifier untuk Kasus Filtering SMS Spam” dengan baik dan lancar.

Penulisan laporan ini merupakan kelengkapan dan pemenuhan dari salah satu syarat dalam memperoleh gelar Sarjana Komputer dari Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana. Selain itu, penulisan laporan Tugas Akhir ini juga bertujuan untuk melatih mahasiswa agar dapat menghasilkan suatu karya yang dapat dipertanggungjawabkan secara ilmiah dan bermanfaat bagi penggunanya.

Dalam penulisan penelitian dan laporan Tugas Akhir ini, penulis banyak menerima bimbingan, saran, dan masukan dari berbagai pihak, baik secara langsung maupun secara tidak langsung. Untuk itu dengan segala kerendahan hati, pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada :

1. Bapak Budi Susanto, S. Kom., M. T. selaku dosen pembimbing I yang selalu memberikan masukan, saran, dan dorongan yang membangun dalam penulis menyelesaikan penelitian dan laporan Tugas Akhir.
2. Ibu Rosa Delima, S. Kom., M. Kom. selaku dosen pembimbing II yang juga selalu memberikan masukan, saran, dan dukungan dalam penulis menyelesaikan penelitian dan laporan Tugas Akhir.
3. Keluarga Tri Santo – Puji Lestari yang selalu memberikan doa dan dukungan serta memfasilitasi penulis dalam menyelesaikan Tugas Akhir.
4. Rekan-rekan penulis yang telah memberikan bantuan, arahan, dan saran sehingga pengerjaan Tugas Akhir ini menjadi lancar dan lebih baik.
5. Pihak lain yang tidak dapat penulis sebutkan satu per satu, sehingga Tugas Akhir ini dapat terselesaikan dengan baik.

Penulis menyadari bahwa penelitian dan laporan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari pembaca sekalian, sehingga suatu saat nanti penulis dapat memberikan karya yang lebih baik lagi.

Akhir kata penulis memohon maaf bila ada kesalahan dalam penyusunan laporan maupun sewaktu penulis melakukan penelitian Tugas Akhir. Semoga penelitian dan laporan Tugas Akhir ini dapat berguna bagi kita semua.

Yogyakarta, Desember 2012

Penulis

© UKDW

INTISARI

Implementasi Naïve Bayesian Classifier untuk Kasus Filtering SMS Spam

Pada tahun 2011, peredaran SMS spam di Indonesia sangatlah marak. SMS tersebut dapat berisi tentang promosi suatu produk yang sering kali tidak dibutuhkan oleh penerima hingga pesan yang berisi tentang penipuan. Hal tersebut merupakan masalah yang kurang diperhatikan di Indonesia. Namun spam telah menjadi topik yang sangat umum di negara lain. Spam dapat meresahkan masyarakat. Dengan SMS berisi penipuan, masyarakat berpotensi menjadi korban dari pengirim SMS tersebut.

Untuk mengatasi permasalahan tersebut, diperlukan suatu sistem yang dapat mengenali SMS spam sehingga SMS dapat dialihkan atau ditandai sebelum sampai pada pengguna. Pada penelitian ini, akan dibangun sebuah sistem yang mengimplementasikan Naïve Bayesian Classifier untuk mengklasifikasi SMS spam, sehingga pengguna dapat mengenali SMS spam tersebut.

Hasil dari penelitian ini, sistem yang dibangun dapat mengklasifikasikan SMS ke dalam kategori spam dan bukan spam. Naïve Bayesian Classifier dapat diimplementasikan efektif untuk kasus filtering SMS spam. Penggunaan text preprocessing atau pengolahan teks yang tepat dapat meningkatkan kinerja sistem klasifikasi ini.

Kata Kunci : Naïve Bayesian, Spam, SMS, Feature Selection.

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN JUDUL.....	ii
PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
INTISARI.....	viii
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
BAB 1	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	2
1.5 Metode/Pendekatan	2
1.6 Sistematika Penulisan.....	3
BAB 2	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori	6
2.2.1 <i>Filtering</i> Spam	6
2.2.2 Teorema Bayes.....	8
2.2.3 <i>Naïve Bayesian Classifier</i>	10
2.2.3.1 Pelatihan Sistem.....	13
2.2.3.2 Penerapan Teorema Bayes	14
BAB 3	20
3.1 Spesifikasi Sistem.....	20
3.2 Rancangan Arsitektur Sistem	21
3.3 Korpora.....	22

3.4	Rancangan Basis Data	22
3.4.1	Kamus Data.....	22
3.5	Diagram Use-Case.....	25
3.6	Rancangan Alur Kerja Sistem	26
3.6.1	Alur Kerja Filtering SMS Spam.....	26
3.6.2	Alur Kerja Proses Pelatihan Sistem	27
3.6.3	Alur Kerja Proses <i>Tokenization</i>	27
3.6.4	Alur Kerja Proses <i>Stopword Removal</i>	28
3.6.5	Alur Kerja Proses <i>Feature Selection</i>	29
3.6.6	Alur Kerja Proses Perhitungan Naïve Bayes	30
3.7	Rancangan Antarmuka	32
3.8	Rancangan Pengujian Sistem	34
BAB 4	36
IMPLEMENTASI DAN ANALISIS SISTEM.....		36
4.1	Implementasi Sistem	36
4.1.1	Implementasi Proses <i>Tokenization</i>	36
4.1.2	Implementasi Proses <i>Stopword Removal</i>	38
4.1.3	Implementasi Proses Pelatihan Sistem.....	39
4.1.4	Implementasi Proses <i>Feature Selection</i>	40
4.1.5	Implementasi Proses Naïve Bayes	42
4.2	Implementasi Antarmuka	43
4.2.1	Implementasi Antarmuka Daftar Percakapan	43
4.2.2	Implementasi Antarmuka Percakapan.....	44
4.2.3	Implementasi Antarmuka Daftar SMS Spam.....	46
4.3	Pengujian Sistem	47
4.4	Analisis Sistem	47
BAB 5	52
KESIMPULAN DAN SARAN.....		52
5.1	Kesimpulan.....	52
5.2	Saran.....	52
DAFTAR PUSTAKA	54

© UKDW

DAFTAR GAMBAR

Gambar 2.1 Algoritma Feature Selection.....	13
Gambar 2.2 Algoritma Pelatihan dan Penerapan Klasifikasi.....	17
Gambar 3.1 Rancangan Arsitektur Sistem.....	21
Gambar 3.2 Use Case Diagram.....	25
Gambar 3.3 Flowchart Filtering SMS Spam.....	26
Gambar 3.4 Flowchart Pelatihan Sistem.....	27
Gambar 3.5 Flowchart Proses <i>Tokenization</i>	28
Gambar 3.6 Flowchart <i>Stopword Removal</i>	29
Gambar 3.7 Flowchart <i>Feature Selection</i>	30
Gambar 3.8 Flowchart Perhitungan Naïve Bayes	31
Gambar 3.9 Rancangan Antarmuka Daftar Percakapan	32
Gambar 3.10 Rancangan Antarmuka Percakapan.....	32
Gambar 3.11 Rancangan Antarmuka Penulisan Pesan Baru	33
Gambar 3.12 Rancangan Antarmuka Daftar SMS Spam.....	34
Gambar 4.1 Antarmuka Daftar Percakapan	43
Gambar 4.2 Antarmuka Percakapan dengan SMS Spam.....	44
Gambar 4.3 Antarmuka Percakapan tanpa SMS Spam.....	45
Gambar 4.4 Antarmuka Daftar SMS Spam	46
Gambar 4.5 Grafik <i>Macro-averaged F₁ Score</i> terhadap Jumlah <i>Feature</i>	50

DAFTAR TABEL

Tabel 2.1 Contoh Dokumen yang Digunakan pada Tahap Training	14
Tabel 2.2 Contoh Data untuk Menerapkan Rumus.....	18
Tabel 3.1 Tabel <i>normalizations</i>	22
Tabel 3.2 Tabel <i>features</i>	23
Tabel 3.3 Tabel <i>categories</i>	24
Tabel 3.4 Tabel <i>conversations</i>	24
Tabel 3.5 Tabel <i>messages</i>	24
Tabel 3.6 Tabel <i>stoplists</i>	25
Tabel 3.7 Variabel untuk Menghitung <i>Precision</i> dan <i>Recall</i>	34
Tabel 4.1 Pseudo-code Tokenization	36
Tabel 4.2 <i>Pseudo-code</i> Pelatihan Sistem	39
Tabel 4.3 <i>Pseudo-code</i> Feature Selection	40
Tabel 4.4 <i>Pseudo-code</i> Kalkulasi Naive Bayes	42
Tabel 4.5 Hasil Pengujian Sistem	47
Tabel 4.6 Hasil Pengujian dengan <i>Feature Selection</i> 100% dari Data Pelatihan .	48
Tabel 4.7 Hasil Perhitungan <i>Precision</i> dan <i>Recall</i>	49
Tabel 4.8 Hasil Perhitungan <i>Macro-averaged F₁ Score</i>	50

INTISARI

Implementasi Naïve Bayesian Classifier untuk Kasus Filtering SMS Spam

Pada tahun 2011, peredaran SMS spam di Indonesia sangatlah marak. SMS tersebut dapat berisi tentang promosi suatu produk yang sering kali tidak dibutuhkan oleh penerima hingga pesan yang berisi tentang penipuan. Hal tersebut merupakan masalah yang kurang diperhatikan di Indonesia. Namun spam telah menjadi topik yang sangat umum di negara lain. Spam dapat meresahkan masyarakat. Dengan SMS berisi penipuan, masyarakat berpotensi menjadi korban dari pengirim SMS tersebut.

Untuk mengatasi permasalahan tersebut, diperlukan suatu sistem yang dapat mengenali SMS spam sehingga SMS dapat dialihkan atau ditandai sebelum sampai pada pengguna. Pada penelitian ini, akan dibangun sebuah sistem yang mengimplementasikan Naïve Bayesian Classifier untuk mengklasifikasi SMS spam, sehingga pengguna dapat mengenali SMS spam tersebut.

Hasil dari penelitian ini, sistem yang dibangun dapat mengklasifikasikan SMS ke dalam kategori spam dan bukan spam. Naïve Bayesian Classifier dapat diimplementasikan efektif untuk kasus filtering SMS spam. Penggunaan text preprocessing atau pengolahan teks yang tepat dapat meningkatkan kinerja sistem klasifikasi ini.

Kata Kunci : Naïve Bayesian, Spam, SMS, Feature Selection.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Hingga akhir tahun 2011, jumlah pelanggan telekomunikasi selular di Indonesia mencapai 250 juta pelanggan. Angka tersebut melampaui jumlah penduduk Indonesia yang mencapai sekitar 240 juta penduduk . Jumlah pelanggan telekomunikasi selular tersebut diprediksi akan terus mengalami peningkatan di masa yang akan datang. Namun di tengah meningkatnya jumlah pelanggan telekomunikasi selular tersebut SMS spam mulai bermunculan.

SMS spam berisi informasi yang tidak dikehendaki oleh penerima pesan. SMS spam dikirim oleh satu pengirim ke banyak nomor yang didapatkan secara acak. SMS spam tersebut biasanya berisi informasi mengenai penipuan, penawaran produk dan informasi tidak penting lainnya. Dalam sebagian kasus pengirim menggunakannya untuk menipu penerima pesan tersebut. Penipuan tersebut telah meresahkan masyarakat pelanggan telekomunikasi selular di Indonesia. Masyarakat semakin resah karena lambannya kinerja aparat penegak hukum untuk mengungkap kasus tersebut. SMS spam yang berisi penawaran produk maupun informasi tidak penting lainnya juga cukup mengganggu pelanggan. Oleh karena itu, kasus SMS spam ini menjadi masalah yang penting untuk diteliti.

Kasus SMS spam dapat diatasi dengan mengklasifikasikan pesan singkat ke dalam kategori spam. Pengklasifikasian tersebut dapat dilakukan secara otomatis oleh sebuah aplikasi dengan mengimplementasikan salah satu metode pengklasifikasian yaitu metode klasifikasi Naïve Bayesian. Metode tersebut membutuhkan pelatihan sistem sebelum dapat dipakai untuk pengklasifikasian. Dalam penelitian ini, pelatihan sistem dibuat di sisi server dan pengklasifikasian dilakukan seluruhnya di aplikasi klien.

Besar harapan penulis, hasil dari penelitian ini dapat membantu pelanggan telekomunikasi selular agar dapat terhindar dari SMS spam. Dengan terhindar dari

SMS spam, pelanggan berarti juga terhindar dari penipuan yang disebarkan melalui SMS spam tersebut.

1.2 Perumusan Masalah

Berdasarkan uraian latar belakang di atas, maka masalah yang akan diteliti dirumuskan sebagai berikut :

- *Text preprocessing* seperti apa yang efektif untuk diterapkan pada pesan singkat di ponsel ?
- Seberapa akurat metode *Naïve Bayesian Classifier* dalam kasus *filtering* SMS spam pada perangkat mobile ?

1.3 Batasan Masalah

Pada permasalahan yang akan diteliti, batasan masalah yang digunakan adalah :

- Kinerja sistem diuji terbatas pada korpora yang berhasil dikumpulkan
- Kinerja fitur pelatihan sistem tidak diuji dalam penelitian ini
- Web service digunakan hanya untuk fitur pelatihan sistem, proses klasifikasi dilakukan di aplikasi klien

1.4 Tujuan Penelitian

Tujuan penelitian dari tugas akhir ini adalah membuat aplikasi *mobile* yang dapat mendeteksi SMS spam menggunakan metode klasifikasi *Naïve Bayesian* pada *platform* Android dan mengetahui kinerja dari metode tersebut sehingga dapat membantu masyarakat agar terhindar dari SMS tersebut.

1.5 Metode/Pendekatan

- Pembuatan Korpora

Hal ini dilakukan untuk membuat korpora yang dibutuhkan untuk melatih dan menguji sistem. Korpora tersebut dikumpulkan dari SMS pribadi milik rekan-rekan penulis.

- Perancangan Program

Perancangan program dilakukan berdasarkan data yang diperoleh dan konsep yang didapat dari studi literatur yang meliputi :

- Perancangan basis data, bahasa pemrograman dan platform yang dibutuhkan dalam pembangunan program.
- Perancangan proses-proses yang terjadi di dalam sistem.
- Perancangan antarmuka dan gambaran kerja dari program yang akan dibangun.

- Pembuatan Program

Pembuatan *prototype* program dilakukan berdasarkan perancangan yang telah dibuat.

- Pengujian Program

Pengujian dilakukan menggunakan korpora yang telah terkumpul. Korpora yang digunakan untuk menguji berbeda dengan korpora yang digunakan untuk melatih sistem.

1.6 Sistematika Penulisan

Sistematika penulisan pada laporan ini secara garis besar dibagi menjadi tiga bagian utama, yaitu bagian awal, bagian utama dan bagian akhir. Pada bagian awal, laporan memuat beberapa poin awal, yaitu halaman sampul depan, sampul dalam, pernyataan keaslian skripsi, halaman persetujuan, halaman pengesahan, ucapan terima kasih, intisari, daftar isi, daftar gambar, dan daftar tabel.

Pada bagian utama, laporan terdiri dari lima bagian, yaitu pendahuluan, tinjauan pustaka, analisis dan perancangan, implementasi serta analisis sistem, dan kesimpulan dan saran. Rincian kelima bagian tersebut adalah sebagai berikut :

- Bagian pendahuluan berisi enam poin penjabaran, yaitu latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, metode/pendekatan dan sistematika penulisan.
- Bagian tinjauan pustaka dibagi menjadi dua bagian utama, yaitu tinjauan pustaka dan landasan teori. Tinjauan pustaka menguraikan berbagai teori yang didapatkan dari berbagai sumber pustaka yang digunakan untuk

penyusunan laporan ini. Landasan teori memuat uraian konsep dan teori yang dipakai untuk memecahkan permasalahan dalam penelitian ini.

- Bagian analisis dan perancangan mencakup analisis teori-teori yang digunakan dan bagaimana menterjemahkannya ke dalam sistem yang hendak dibuat. Bagian ini memuat bahan dan alat yang dipakai dalam implementasi, variabel yang digunakan, data yang dikumpulkan, penjelasan algoritma dan arsitektur sistem, dan uraian mengenai perancangan dan simulasi sistem.
- Bagian implementasi dan analisis sistem memuat hasil dan analisis dari implementasi sistem. Hasil implementasi disajikan dalam bentuk daftar, table, grafik, foto atau bentuk lain dan ditempatkan sedekat-dekatnya dengan pembahasan sehingga lebih mudah dipahami. Pembahasan tentang hasil yang diperoleh berupa penjelasan teoritis, baik secara kualitatif, kuantitatif atau secara statistis.
- Bagian kesimpulan dan saran berisi pernyataan singkat dan tepat yang dijabarkan dari hasil analisis implementasi. Bagian ini juga berisi mengenai saran-saran yang diberikan untuk penelitian yang dilakukan di masa mendatang. Saran-saran tersebut memuat metode atau teknik pengembangan yang belum digunakan di dalam penelitian namun dirasa akan memperbaiki kinerja sistem.

Pada bagian akhir, laporan memuat penjabaran daftar pustaka dan bagian lampiran.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dalam penelitian ini, Naïve Bayesian *classifier* telah berhasil diimplementasikan untuk *filtering* SMS spam. berdasarkan hasil pengujian dan analisis yang telah dilakukan, maka dapat disimpulkan bahwa :

1. Sistem mampu melakukan *filtering* terhadap SMS spam pada perangkat mobile Android dengan *Macro-averaged F₁ Score* sebesar 0,82. *Macro-averaged F₁ Score* tersebut menunjukkan tingkat akurasi yang dicapai oleh sistem.
2. *Text preprocessing* yang optimal dan diterapkan pada penelitian ini adalah :
 - a. *Tokenization* dengan *delimiter* berupa spasi, titik dan koma
 - b. *Stopword Removal*
 - c. *Feature Selection* dengan parameter 90%

5.2 Saran

Saran-saran yang dapat digunakan dalam pengembangan aplikasi ke depannya antara lain :

1. Dalam pengembangan selanjutnya, sebaiknya data pelatihan perlu ditambah agar kinerja sistem dapat ditingkatkan. Naïve Bayesian cenderung membutuhkan data pelatihan lebih banyak agar dapat menghasilkan kinerja yang lebih baik lagi.
2. Banyak hal yang dapat diteliti lebih lanjut lagi dalam topik ini. Setiap proses seperti *tokenization*, *normalization*, *stopword removal*, atau *feature selection* mempunyai algoritma yang beragam. Dari algoritma yang beragam tersebut dapat diteliti lebih lanjut lagi algoritma manakah yang paling tepat untuk kasus *filtering* SMS spam ini sehingga sistem yang

dibangun dalam penelitian ini dapat dikembangkan dan dioptimalkan kinerjanya.

© UKDW

DAFTAR PUSTAKA

- Cormack, G.V. (2008). Email Spam Filtering: a Systematic Review. Massachusetts : Now Publishers Inc.
- Cormack, G.V., Hidalgo, J.M.G, & Sanz, E.P. (2007). Spam Filtering for Short Messages. New York: ACM
- Dewi, J. (2005). Filtering Spam Dengan Algoritma Bayesian Combination. Yogyakarta : Universitas Kristen Duta Wacana.
- Elhaj, M.M.K. (2009). Combating SPAM.
<http://www.cert.sd/images/stories/Combating%20SPAM.pdf> diakses 2 Oktober 2012
- Feldman, R., Sanger, J. (2007). The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data. Cambridge : Cambridge University Press
- Guido Schryen. (2007). Anti-Spam Measures Analysis and Design. Berlin : Springer
- Hidalgo, J.M.G., Bringas, G.C., Sanz, E.P, & Garcia, F.C. (2006). Content Based SMS Spam Filtering. New York: ACM
- Kagstorm, J. (2005). Improving Naïve Bayesian Spam Filtering. Sundsvall : Mid Sweden University
- Konchady, M. (2006). Text Mining Application Programming. Massachussets : Charles River Media
- Manning, C. D., Raghavan, P., & Schutze, H. (2008). Introduction to Information Retrieval. Cambridge : Cambridge University Press.
- Robinson, G. (2002). Spam Detection. <http://radio-weblogs.com/0101454/stories/2002/09/16/spamDetection.html> diakses tanggal 20 Juli 2012
- Segaran, T. (2007). Programming Collective Intelligence. California : O'Reily Media