DETEKSI SPAM PADA EMAIL MENGGUNAKAN *INFORMATION GAIN* DAN *ALGORITMA NAÏVE BAYES*

PROPOSAL TUGAS AKHIR

Disusun oleh :

Rendi Wahyu putra

80201090108

Untuk Pesyaratan Penelitian Dan Penulisan Tugas Akhir

Sebagai Akhir Proses Studi Strata 1

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS DINAMIKA BANGSA

JAMBI

2022

IDENTITAS PROPOSAL PENELITIAN

Judul Proposal	:	Deteksi Spam Pada Email Menggunakan Information gain Dan Algoritma Naïve Bayes
Program Studi	:	Teknik Inforamatika
Jenjang Pendidikan	:	Strata 1 (S1)
Peneliti	:	

a.	Nama Lengkap	:	Rendi Wahyu Putra
b.	NIM	:	8020190108
c.	Jenis Kelamin	:	Laki-laki
d.	Tempat/Tgl Lahir	:	Padang/16 januari 2002
e.	Alamat	:	Jl. F. Silean, Rt.25, Kec. Kebun Handil
f.	No. Telepon	:	0895401122244
g.	Email	:	rendiwahyuputra.9e@gmail.com

DAFTAR HALAMAN

BAB I
PENDAHULUAN

1.1 LATAR BELAKANG

Dewasa ini, penggunaan email (*electronic mail*/surat elektronik) sebagai salah satu saran komunikasi modern telah menjadi sesuatu yang sangat umum. Terbukti luasnya jangkauan penggunaan email itu sendiri. Email saat ini tidak hanya digunakan sebagai sarana komunikasi antar individu, juga digunakan sebagai sarana komunikasi antar instansi. Email memiliki keunggulan dibandingkan dengan surat konvensional terutama dalam kaitannya dengan efesiensi waktu dan biaya. Tetapi hal tersebutlah menjadi kelemahan email itu sendiri. Dikarenakan mengirim email tidak memerlukan biaya dan sangat cepat, seringkali hal ini dimanfaatkan oleh para spammer (orang yang mengirimkan email spam/sampah) untuk dengan mudah melakukan aktivitasnya.

Email spam atau juga dikenal sebagai pesan sampah adalah pesan massal yang tidak diminta kemudian dikirim melalui email. Penggunaan spam telah populer sejak awal 1990-an dan merupakan sebuah masalah yang terus dihadapi oleh sebagian besar pengguna email. Pengirim spam biasanya mengirim email ke jutaan alamat email, dengan harapan bahwa sebagian kecil akan membalas atau merespon dengan pesan tersebut[1].

Akan tetapi, dengan berkembangnya serta banyaknya layanan dan teknologi anti-spam, jumlah pesan spam tetap bertambah secara pesat[2]. Hal ini akan menjadi sebuah ancaman bukan hanya di dunia internet saja, tetapi untuk masyarakat juga akan berpengaruh. Sebagai contoh, ketika seorang pengguna menerima email spam dalam jumlah yang sangat besar, peluang pengguna lupa untuk membaca email yang bukan merupakan spam semakin meningkat. Akibatnya, banyak pembaca email harus menghabiskan waktunya hanya untuk menghapus pesan-pesan yang tidak diinginkan. Email spam juga menghabiskan

uang pengguna dalam pemborosan bandwidth atau kuota, serta memungkinkan pengguna untuk melihat konten yang tidak diinginkan. Beberapa tahun terakhir, banyak upaya yang diberikan untuk memblokir email spam.

Spam filtering dapat mengontrol masalah ini dalam banyak cara. Dalam sebuah filtering, beberapa email spam tidak ditandai sebagai spam karena email filtering tersebut tidak mendeteksi bahwa email itu adalah email spam. Berbagai algoritma pembelajaran mesin telah digunakan untuk menyaring email spam. Salah satunya adalah metode Naive Bayes yang populer di dalam kebutuhan komersil dan penyaringan spam yang bersifat open-source[3]. Karena metode ini cukup sederhana, sehingga dapat memudahkan untuk mengimplementasi dan membutuhkan waktu yang cukup singkat untuk memproses data latih dan evaluasi yang cepat untuk menyaring email spam. Penyaringan membutuhkan data latih yang didapatkan dari kumpulat set pesan spam dan nonspam yang telah ada sebelumnya. Dalam data tersebut akan dilacak tiap kata yang muncul hanya dalam pesan spam, pesan non-spam, ataupun keduanya. Naive Bayes juga dapat digunakan pada dataset lain yang tiap datanya memiliki fitur dan atribut yang berbeda

Data mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual[4]. *Naïve bayes* adalah bagian dalam data mining termasuk dalam metode klasifikasi dan merupakan kelompok algoritma *Decision Tree*. Oleh karenanya penelitian ini bertujuan meningkatkan performa deteksi serangan *Brute Force* dengan menggunakan teknik seleksi fitur. Pada penelitian ini *Information Gain* digunakan sebagai seleksi fitur dan berfungsi sebagai meningkatkan fungsi *Naïve Bayes. Naïve Bayes* digunakan metode klasifakasi yang akan ditingkatkan performanya. Hal inilah yang melatar belakangi penulis untuk melakukan penelitian dengan judul **“DETEKSI SPAM PADA EMAIL MENGGUNAKAN INFORMATION GAIN DAN ALGORITMA NAÏVE BAYES’’**

1.2 RUMUSAN MASALAH

Rumusan permasalahan yang akan diselesaikan dalam penelitian ini yaitu :

a. Bagaimana mendeteksi spam pada email di Dataset Spam Assassin
b. Bagaimana menerapakan Information Gain dan Naïve Bayes untuk mengenali email spam pada dataset Spam Assassin

1.3 BATASAN MASALAH

Batasan masalah yang akan di bahas pada penelitian ini sebagai berikut:

a. Menggunakan dataset Spam Assassin
b. Mengklasifikasikan email spam menggunakan Information Gain dan Naïve bayes

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Tujuan dari penelitian yang dilakukan adalah :

a. Menerapkan metode Information Gain dan Naïve Bayes untuk mengelompokkan email spam dan bukan email spam.
b. Mengklasifikasikan email spam pada data Spam Assassin dengan menggunakan metode kaslifikasi.

1.4.2 Manfaat penenelitian

Adapun manfaat dari penelitian ini adalah :

a. Dapat mengetahui pola email spam pada dataset Spam Assassin
b. Dapat mengetahui akurasi serangan dengan menggunakan seleksi fitur Information gain dan algoritma Naïve Bayes.

BAB II
LANDASAN TEORI

2.1 DATA MINING

Menurut Gartner Group,data mining adalah proses menemukan hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilah-milah sebagian besardata yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola seperti teknik statistik dan matematika. Data mining merupakan gabungan dari beberapa disiplin ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar[5]. Data Mining disebut juga sebagai *Knowledge Discovery in Database (KDD)*, didefinisakan sebagai ekstraksi informasi potensial , implisit dan tidak dikenal dari sekumpulan data.

Ada beberapa macam pendekatan yang berbeda yang diklasifikasikan sebagai teknik pencarian informasi/pengetahuan dalam *KDD.* Ada pendekatan kuantitatif, seperti pendekatan probabilistik seperti logika induktif, pencarian pola, dan analisis pohon keputusan. Pendekatan yang lain meliputi deviasi, analisis kecenderungan, algoritma genetik, jaringan saraf tiruan, dan pendekatan campuran dua atau lebih dari beberapa pendekatan yang ada[6].

Pada dasarnya ada enam elemen yang paling esensial dalam Teknik pencarian informasi/pengetahuan dalam *KDD* yaitu:

1. Mengerjakan sejumlah besar data.
2. Diperlukan efesiensi berkaitan dengan volume data.
3. Mengutamakan ketetapan/keakuratan.
4. Membutuhkan pemakaian bahasa tingkat tinggi.
5. Menggunakan beberapa bentuk dari pembelajaran otomatis.
6. Menghasilkan hasil yang menarik.

2.1.1 Definisi Data Mining

Berdasarkan defenisi-defenisi yang telah disampaikan, hal penting yang terkait dengan Data Mining adalah:

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang akan mungkin memberikan indikasi yang bermanfaat.

2.1.2 Teknik Data Mining

Teknik-teknik dalam data mining adalah sebagai berikut[4]:

1. *Classification*

   Klasifikasi adalah teknik yang paling umum diterapkan pada data mining. Pendekatan ini sering menggunakan keputusan pohon (decision tree) atau neural network berbasis algoritma klasifikasi. Proses klasifikasi data melibatkan learning dan klasifikasi. Dalam belajar (learning) data pelatihan (training) dianalisis dengan alogaritma klasifikasi. Dalam klasifikasi pengujian data dilakukan dengan menggunakan perkiraan akurasi dari aturan klasifikasi. Jika akurasi bisa diterima, maka aturan dapat diterapkan untuk data baru. Salah satu contoh yang mudah dan populer adalah dengan decision tree yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi. Decision tree adalah model prediksi menggunakan struktur pohon atau struktur berhirarki.

   Decision tree adalah struktur flowchart yang menyerupai tree (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas. Alur pada decision tree di telusuri dari simpul akar ke simpul daun yang memegang

prediksi kelas untuk contoh tersebut. Decision tree mudah untuk dikonversi ke aturan klasifikasi (classification rules).

2. *Clustering*

   Clustering bisa dikatakan sebagai identifikasi kelas objek yang memiliki kemiripan. Dengan menggunakan teknik clustering kita bisa lebih lanjut mengidentifikasi kepadatan dan jarak daerah dalam objek ruang dan dapat menemukan secara keseluruhan pola distribusi dan korelasi antara atribut. Pendekatan klasifikasi secara efektif juga dapat digunakan untuk membedakan kelompok atau kelas objek.

3. *Association rule*

   Digunakan untuk mengenali kelakuan dari kejadian-kejadian khusus atau proses dimana link asosiasi muncul pada setiap kejadian. Contoh dari aturan assosiatif dari analisa pembelian di suatu pasar swalayan adalah bisa diketahui berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu.

   Penting tidaknya suatu aturan assosiatif dapat diketahui dengan dua parameter, support yaitu prosentasi kombinasi atribut tersebut dalam basisdata dan confidence yaitu kuatnya hubungan antar atribut dalam aturan asosiatif. Motivasi awal pencarian association rule berasal dari keinginan untuk menganalisa data transaksi supermarket, ditinjau dari perilaku customer dalam membeli produk. Association rule ini menjelaskan seberapa sering suatu produk dibeli secara bersamaan. Sebagai contoh, association rule “beer => diaper (80%)” menunjukkan bahwa empat dari lima customer yang membeli beer juga membeli diaper. Dalam suatu association rule X => Y, X disebut dengan antecedent dan Y disebut dengan consequent rule.

4. *Predication*

   Teknik regresi dapat disesuaikan untuk predikasi. Analisis regresi dapat digunakan untuk model hubungan antara satu atau lebih independent variables dan dependent variables. Dalam data mining independent variabel adalah atribut-atribut yang sudah dikenal dan respon variabel apa yang kita inginkan untuk diprediksi. Akan tetapi, banyak masalah di dunia nyata bukan prediksi yang mudah. Karena itu, teknik kompleks (seperti: logistic regression, decision trees atau pohon keputusan, neural nets atau jaringan syaraf) mungkin akan diperlukan untuk memprediksi nilai. Model yang berjenis sama sering dapat digunakan untuk regresi dan klasifikasi. Misalnya, CART (Classification and Regression Trees) yaitu alogaritma pohon keputusan yang dapat digunakan untuk membangun kedua pohon klasifikasi dan pohon regresi. Jaringan saraf juga dapat menciptakan kedua model klasifikasi dan regresi

5. *Neural Network*

   Jaringan saraf adalah seperangkat unit penghubung input dan output dimana setiap koneksinya memiliki bobot. Selama fase learning, jaringan belajar dengan menyesuaikan bobot sehingga dapat memprediksi kelas yang benar label dari setiap input. Jaringan saraf memiliki kemampuan yang luar biasa untuk memperoleh arti dari data yang rumit atau tidak tepat dan dapat digunakan untuk mengambil pola- pola serta mendeteksi tren yang sangat kompleks untuk diperhatikan baik oleh manusia atau teknik komputer lain. Jaringan saraf sangat baik untuk mengidentifikasi pola atau tren pada data dan sangat cocok untuk melakukan prediksi serta memprediksi kebutuhan.

6. *Decision Tress*

   Decision trees atau pohon keputusan adalah struktur *tree-shaped* yang mewakili set keputusan. Keputusan ini menghasilkan aturan untuk klasifikasi sebuah kumpulan data. Metode pohon keputusan

diantaranya yaitu *Classification and regression trees* (CART) dan *Chi Square Automatic Interaction Detection* (CHAID).

7. *Nearest Neighbor Method*

   Teknik yang mengklasifikasikan setiap *record* dalam sebuah kumpulan data berdasarkan sebuah kombinasi suatu kelas k record yang sama dalam sebuah kumpulan data historis (dimana k lebih besar atau sama dengan 1). Terkadang disebut juga dengan teknik *K-Nearest Neighbor.*

2.1.3 Fungsi Dan Tugas Data Mining

Data mining menganalisis data menggunakan tool untuk menemukan pola dan aturan dalam himpunan data. Perangkat lunak bertugas untuk menmukan pola dengan mengidentifikasi aturan dan fitur pada data. Tool data mining diharapkan mampu mengenal pola ini dalam data dengan input minimal dari user[6].

2.2 SPAM

Spam atau junk mail adalah penyalahgunaan dalam pengiriman berita elektronik untuk menampilkan berita iklan dan keperluan lainnya yang mengakibatkan ketidaknyamanan bagi para pengguna web[7].

Berikut ini definisi dari spam[8].

1. Isi atau konten dari email tidak relevan dengan minat penerima.
2. Penerima tidak dapat menolak datangnya email yang tidak diminta tersebut dengan cara-cara lazim.
3. Dari sisi penerima, pengiriman dan penerimaan pesan tersebut memberikan keuntungan bagi pengirimnya.

Bentuk berita spam yang umum dikenal meliputi: spam pos-el, spam pesan instan, spam usenet news-group, spam mesin pencari informasi web (web search

engine spam), spam blog, spam berita pada telepon genggam, spam forum internet, dan lain lain. Spam ini biasanya datang bertubi-tubi tanpa diminta dan sering kali tidak dikehendaki oleh penerimanya.

Spam terjadi akibat murahnya biaya untuk mengirimkan spam. Biaya untuk mengirimkan satu email sama dengan seribu email, atau bahkan satu juta email[9].

Spam dapat dikategorikan sebagai berikut

1. Junk mail yaitu email yang dikirimkan secara besar-besaran dari suatu perusahaan bisnis, yang sebenarnya tidak kita inginkan.
2. Non-commercial spam, misalnya surat berantai ataucerita humor yang dikirimkan secara masal tanpatujuan komersial tertentu.
3. Pornographic spam yaitu email yangdikirimkansecara masal untuk mengirimkan gambar-gambarpornografi.
4. Virus spam yaitu email yang dikirimkansecaramassal, dan mengandung virus atau Trojans.

2.3 ALGORITMA NAÏVE BAYES

Naive Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Naive Bayes* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. *Naive Bayes* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar[10]. *Naïve Bayes* memiliki kecepatan dan akurasi yangtinggi ketika diaplikasikan pada sebuah data yang besar [11]

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

P(H\X) : Probabilitas hipotesis H berdasarkan kondisi X (posteriori prob)

P(H) : Probabilitas hipotesis H (prior prob)

P(X\H) : Probabilitas X berdasarkan kondisi tersebut

P(X) : Probabilitas dari X

2.4 SELEKSI FITUR

2.4.1 Information Gain

Information Gain adalah jumlah informasi yang disediakan oleh fitur teks untuk kategori teks. Information Gain dihitung dengan berapa banyak term yang digunakan untuk klasifikasi informasi, untuk mengukur pentingannya *lexical teks* untuk diklasifikasi[11]. IG dapat membantu mengurangi noise yang disebabkan oleh fitur-fitur yang tidak relevan. IG mendeteksi fitur-fitur yang paling banyak memiliki informasi berdasarkan kelas tertentu. Penentuan atribut terbaik dilakukan dengan menghitung nilai *entropi* terlebih dahulu. *Entropi* merupakan ukuran ketidakpastian dapat digunakan untuk menyimpulkan distribusi fitur dalam bentuk yang ringkas.

2.4.2 Gain Ratio

Salah satu pembobotan menggunakan GR. GR dapat memperbaiki data yang tidak stabil, cocok untuk data numerik dua kelas, sederhana sehingga komputasi lebih cepat. Untuk menghitung gain ratio diperlukan pemisahan informasi (split).

2.4.3 Corelation-Based (CB)

Algoritma correlation-based termasuk filter method yang menggunakan *heuristik* untuk mengevaluasi kemampuan (merit) subset

fitur. Sebuah hipotesis yang menjadi dasar dari heuristik didasarkan atas pernyataan : "*Good feature subsets contain features highly correlated with the class, yet uncorrelated with each other"*.

2.4.4 CFs-PSO Search

PSO merupakan teknik komputasi yang evolusioner, mudah di implementasi dan melakukan komputasi secara effisien[12]. Pada penelitian ini mengadopsi teknik yang menggunakan teknik Correlation-based Feature Selection (CFS) yang dioptimasi dengan Particle Swarm Optimization (PSO).

2.5 WEKA

WEKA merupakan sebuah system data *Data Mining* yang dikembangkan oleh Universitas Waikato di Selandia Baru mengimplementasikan algoritma Data Mining. *WEKA* menyediakan implementasi dari pembelajaran algoritma yang dapat dengan mudah untuk diterapkan pada dataset. Implementasi tersebut juga mencakup berbagai alat untuk mengubah dataset, *Pre-proccess* dataset, memberikan skema pembelajaran dan menganalisis klasifikasi yang dihasilkan dan kinerjanya tanpa harus menuliskan kode program.

2.6 PENELITIAN SEJENIS

Kajian penelitian sejenis adalah kajian dari pendapat orang lain untuk menjadi perbandingan atau acuan bagi penulisan ilmiah penulis. Penelitian ini dilakukan tidak lepas dari hasil penelitian-penelitian terhadap yang pernah dilakukan sebagai bahan perbandingan atau kajian. Adapun hasil-hasil penelitian yang disajikan perbandingan tidak terlepas dari topik penelitian yaitu Penerapan *Data Mining* dalam menentukan prediksi akurasi. Berikut adalah beberapa penelitian sejenis terdahulu yang dikaji dalam bentuk tabel :

tabel 2. 1 Penelitian sejenis

No	Judul, Penulis dan Tahun	Masalah	Metode	Ringkasan Hasil
1	Klasifikasi Email Spam Dengan Menggunakan Metode Support Vector Machine Dan K-Nearest Neighbor[13]	Akibat dari penggunaan email yang sangat intens dapat menyebabkan dampak positif dan negatif. Hal ini dikarenakan tidak semua orang dapat menggunakan email dengan baik dan diketahui banyak sekali penyalahgunaan email yang berpotensi dapat merugikan perusahaan ataupun individual. Email yang disalahgunakan ini biasa dikenal sebagai spam atau junkmail (email sampah), isi dari email tersebut bisa berupa iklan penjualan produk, penipuan berkedok menang undian atau bahkan virus dan malware. Banyaknya penyalahgunaan email ini menimbulkan	Support Vector Machine Dan K-Nearest Neighbor	a. Metode KNN memberikan hasil performansi klasifikasi terbaik saat k = 3 dengan hasil ketepatan klasifikasi, precision dan recall berturut-turut sebesar 92.28%, 92.3% dan 92.3% serta error sebesar 7.72% dari total 6000 email. b. Metode SVM memberikan hasil klasifikasi terbaik dengan menggunakan kernel linier, nilai yang diberikan untuk ketepatan klasifikasi, precision dan recall berturut-turut sebesar 96.6%, 96.6% dan 96.6% serta error sebesar 3.4% dari total 6000 email. c. SVM dengan kernel linier mampu memberikan hasil klasifikasi yang lebih baik dibandingkan dengan 3- NN.

		kerugian yang cukup besar antara lain dapat meningkatkan data traffic dan menyebabkan kerugian ekonomis yang cukup signifikan, terutama bagi perusahaan.		
2	Deteksi Surel Spam Dan Non Spam Bahasa Indonesia Menggunakan Metode Naïve Bayes[14]	Penggunaan surel yang mudah saat ini banyak sekali dimanfaatkan banyak orang sehingga menimbulkan dampak positif maupun negatif. Surel negatif biasa kita sebut dengan surel spam yang berisi berupa iklan, penipuan, virus dan malware yang berpotensi untuk merugikan orang lain. Masalah tersebut memerlukan penanganan untuk mengatasinya.	Naïve Bayes	Dari hasil penelitian ini telah berhasil dibangun sebuah model untuk melakukan deteksi surel Spam dan non Spam berbahasa Indonesia dan Algoritma Naïve Bayes memiliki akurasi yang sangat baik. Berdasarkan eksperimen yang dilakukan perpaduan fitur N-Gram dengan nilai n=2 dan algoritma Naïve Bayes akurasi yang paling tinggi yaitu 94%
3	Deteksi Pesan Spam pada Forum Daring Menggunakan Metode Naïve Bayes[15]	Forum diskusi daring adalah salah satu layanan edukasi yang disediakan oleh pihak Universitas XYZ yang dapat digunakan oleh dosen dan mahasiswa.	Naïve Bayes	Berdasarkan penelitian yang telah dilakukan untuk menangani permasalahan tersebut, dengan menggunakan data sebanyak 100 pesan yang berasal dari situs forum daring yang dikhususkan untuk

		Forum diskusi ini dijadikan sebagai sarana untuk mengeluarkan pendapat mahasiswa mengenai pelajaran serta sebagai pendataan absensi kehadiran mahasiswa tersebut. Pada prakteknya sering sekali mahasiswa hanya memberikan tanggapan yang seharusnya tidak diunggah pada forum diskusi berupa pesan spam sehingga fungsi utama dari forum daring sebagai media pembelajaran dan sarana untuk mengeluarkan pendapat menjadi terkesampingkan		mahasiswa program studi teknik informatika Universitas XYZ tahun ajaran 2021/2022. Peneliti mendapatkan temuan dimana klasifikasi pesan spam menggunakan algoritma Naive Bayes Classifier (NBC) memperoleh nilai accuracy sistem sebesar 95%, nilai precision sistem sebesar 100%, dan nilai recall sistem sebesar 90%. Penggunaan algoritma NBC dapat dijadikan salah satu alternatif dalam mendeteksi pesan spam pada forum.
4	Deteksi E-Mail Dan Spam Menggunakan Fuzzy Association Rule Mining[16]	Munculnya suatu e-mail komersial yang tidak diharapkan atau yang lebih sering disebut dengan spam sangat menggangu pengguna e-mail karena dapat menambah penggunaan bandwith koneksi internet, serta akan menjadi	logika fuzzy dan association rule	Berdasarkan pengujian yang dilakukan, penggunaan metode Fuzzy Association Rule Mining dalam pendeteksian SPAM dan non SPAM pada e-mail adalah baik, hal ini terlihat pada pengujian dimana nilai accuracy, precision, recall, dan F-Measure cukup

		suatu sampah yang menumpuk sehingga mengurangi kapasitas penyimpanan.		tinggi yang mendekati nilai 1.
5	Klasifikasi Komentar Spam Pada Instagram Berbahasa Indonesia Menggunakan K-NN[17]	ara publik figur, terutama aktor dan artis menggunakan Instagram (IG) sebagai salah satu media sosial berbasis foto untuk mempromosikan kegiatan dan menjalin relasi dengan para penggemarnya. Para penggemar dapat memfollow, melihat foto, dan berkomentar pada setiap status IG artis idolanya. Permasalahannya, banyak sekali komentar spam yang ditulis pada IG yang sampai saat ini belum ada penyelesaiannya, terutama untuk spam berbahasa Indonesia.	K-Nearest Neighbor (K-NN)	a. Algoritma k-NN dapat diimplementasikan dengan sangat baik pada kasus klasifikasi komentar spam pada Instagram berbahasa Indonesia dengan tingkat rata-rata akurasi dari 4 skenario 87.07 %. Hal ini masuk dalam kategori hasil yang sangat baik b. Penelitian ini telah menguji sistem deteksi komentar spam pada IG menggunakan algoritma klasifikasi k-NN dan diperoleh rata-rata akurasi sebesar 87.84 % untuk skenario I (UB NS) dan sebesar 88,4% untuk skenario II (UB S). Terjadi peningkatan sebesar 0.56 % akibat dari penggunaan stemming c. Penelitian ini telah menguji sistem deteksi komentar

				spam pada IG menggunakan algoritma klasifikasi k-NN dan diperoleh rata-rata akurasi sebesar 86.17 % untuk skenario III (B NS) dan sebesar 85.86 % untuk skenario IV (B S). Terjadi penurunan 0.31 % pada dataset seimbang (B).
6	Sistem Deteksi Surel SPAM Dengan DNSBL Dan Support Vector Machine Pada Penyedia Layanan Mail Marketing[18]	Mail marketing merupakan media komunikasi bagi pengguna dan penyedia jasa Internet yang efektif. Banyak perusahaan menggunakan surel sebagai media komunikasi dengan pelanggan untuk memastikan pelanggan tidak tertinggal informasi terbaru serta memberikan penawaran personal pada pelanggan tertentu. Meski demikian, tidak semua surel yang dikirim dapat sampai ke kotak masuk surel seperti yang diharapkan. Ada beberapa faktor	DNSBL Dan Support Vector Machine	Proses deteksi SPAM dilakukan dalam dua tahap, yaitu mengecek alamat domain dengan DNSBL dan mengecek konten surel dengan model yang dikembangkan menggunakan SVM. Model yang dikembangkan telah divalidasi dengan 10-fold cross validation dan memiliki rata-rata akurasi 97.54%, sehingga cukup valid untuk digunakan pada SPAM detection and prevention system.

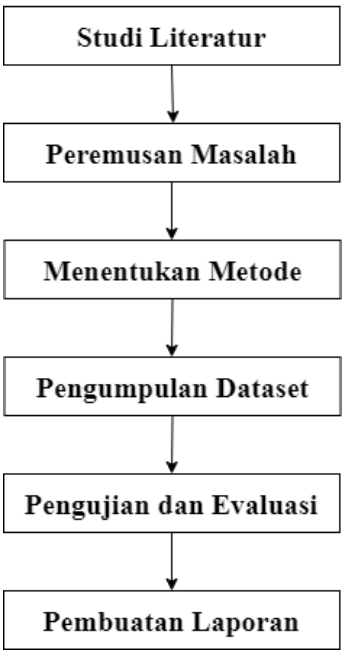
		yang mempengaruhi hal tersebut, antara lain karena konten yang tidak sesuai kaidah penulisan dan cenderung memiliki signature SPAM, alamat tujuan surel yang tidak valid, domain pengguna yang terdaftar dalam blacklist dan lain sebagainya. Penyedia layanan mail marketing seperti MTarget dan Mailchimp harus memastikan email yang dikirim oleh pelanggannya tidak berpotensi menjadi SPAM, karena dapat berdampak seluruh layanan mail marketingnya akan masuk dalam daftar hitam dan tujuan promosi tidak tercapai.		

Dari beberapa penelitian pada tabel 2.1 dapat disimpulkan bahwa metode *Naïve Bayes* memiliki akurasi yang baik dalam menganalisis email spam. Maka dari itu dalam penelitian ini penulis menggunakan metode *Naïve Bayes* untuk menganalisis email spam. Penelitian-penelitian yang telah dilakukan sebelumnya ini akan menjadi acuan dalam melakukan penelitian.

BAB III
METODOLOGI PENELITIAN

3.1 KERANGKA KERJA PENELITIAN

Untuk membantu dalam penyusunan penelitian ini, maka perlu adanya kerangka kerja (framework) yang jelas tahapan-tahapannya. Kerangka kerja ini merupakan langkah- langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas. Adapun kerangka kerja penelitian yang akan digunakan dapat dilihat pada gambar 3.1:



Gambar 3. 1 Kerangka Kerja Penelitian

Berdasarkan kerangka kerja penelitian yang telah di gambarkan pada gambar 3.1, maka dapat di uraikan pembahasan masing masing tahapan dalam penelitian yaitu:

1. **Studi Literatur**

   Mempelajari literatur-literatur yang dapat mencapai tujuan penelitian, pada tahap ini penulis mengumpulkan bahan bacaan yang

bersumber dari buku-buku perpustakaan Universitas Dinamika Bangsa Studi Literatur Perumusan Masalah Menentukan Metode Pengumpulan Dataset Pengujian dan Evaluasi Pembuatan Laporan Jambi serta sumber internet dan ditemukanlah penelitian yang memuat tentang email spam dan dalam penelitian ini akan menjadi referensi untuk pengujian email spam.

2. **Perumusan Masalah**

Pada tahap ini penulis merumuskan masalah yang dirumuskan dalam penelitian ini adalah bagaimana penerapan metode Naïve Bayes dalam mendeteksi email spam, dan tujuan yang akan dicapai dalam penelitian ini adalah ingin mengetahui berapa besar tingkat akurasi metode Naive Bayes dalam mendeteksi email spam.

3. **Menentukan Metode**

Pada tahap ini, dilakukan proses klasifikasi dan menggunakan alat bantu (tools) WEKA terhadap data email spam, selanjutnya dilakukan analisa data dari dataset. Data yang telah didapatkan dari dataset Spam Assassin akan dianalisis dengan menggunakan metode Naive Bayes sehingga didapatkan hasil akurasi dari email spam.

4. **Pengumpulan Dataset**

Dalam tahapan ini penulis mengambil, mengolah dan menganalisis data dari dataset CICIDS-2017 menggunakan metode Naive Bayes.

5. **Pengujian dan Evaluasi**

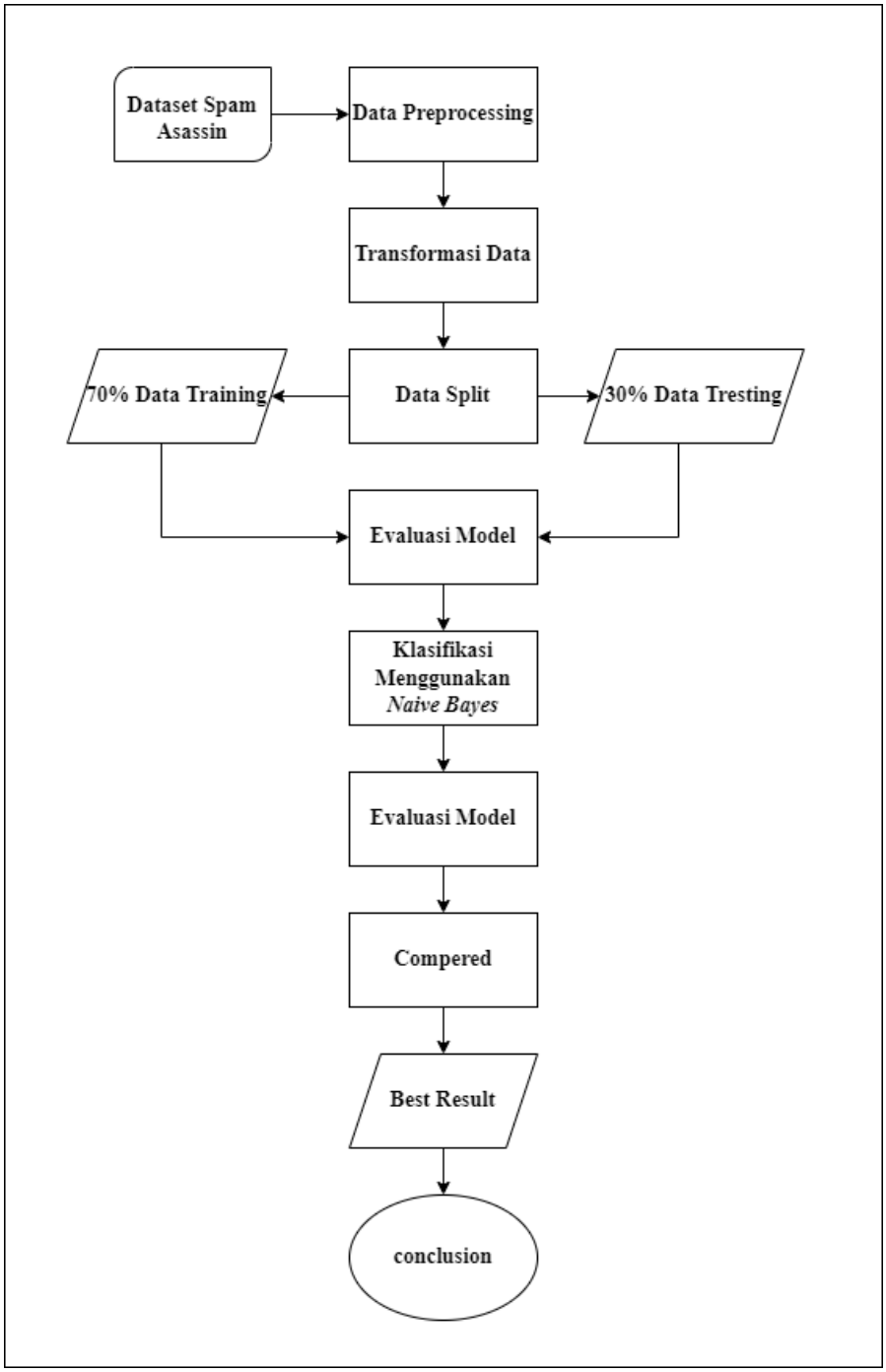
Pada tahapan ini penulis menguji pembuatan Naïve Bayes. Hasil yang didapat bisa digunakan untuk mengetahui berapa besar tingkat akurasi metode Naive Bayes dalam mendeteksi email spam.

6. **Pembuatan Laporan**

Setelah semua tahapan penelitian dilakukan, maka akan dibuat laporan sebagai dokumentasi penelitian agar dapat dimanfaatkan pada waktu yang akan datang baik oleh peneliti sendiri maupun peneliti lainnya.

3.2 ALUR EKSPERIMEN

Untuk membantu dalam pembuatan eksperimen ,maka diperlukan adanya alur eksperimen agar membantu dalam melakukan pengujian dan evaluasi Adapun alur eksperimen yang akan digunakan dapat dilihat pada gambar 3.2



Gambar 3. 2 Alur Eksperimen

1. *Dataset*

   Dataset yang digunakan dalam penelitian ini adalah spam assassin yang hanya terdiri dari email spam dan bukan email spam.

2. *Preprocessing*

   proses yang mengubah data *Dataset* (*Spam Assassin*) ke dalam bentuk yang lebih mudah dipahami.

3. *Transformasi Dataset*

   Dilakukan dengan tujuan utama untuk mengubah skala pengukuran *spam assassin* menjadi bentuk lain sehingga data dapat memenuhi asumsi-asumsi yang mendasari analisis ragam.

4. *Split Data*

   Split Dataset termasuk dalam tahapan preprosesing untuk mendapatkan *dataset spam assassin* yang proporsional.

   a) Data Training

   Data training nantinya akan digunakan untuk melatih algoritma dalam mencari model naive bayes.Data yang di ambil dari dataset untuk data training adalah 70%.

   b) Data Testing

   Data testing akan dipakai untuk menguji dan mengetahui performa naive bayes.Sisa dari dataset akan di gunakan sebagai data testing yaitu 30%.

5. Seleksi Fitur Dengan *Infomartion Gain*

   Menggunakan Information gain untuk menentukan bobot/ranking setiap atribut dan mengambil data dari 100 data yang mempunyai bobot/ranking tertinggi.

6. Klasifikasi Menggunakan *Naïve Bayes*

   Diagram alir dari Naïve bayes dimulai dari membaca data training dengan melihat pola serangan dari brute force yang dimana hasil dari pembacaan pola serangan tersebut merupakan peluang dari 20 setiap pola serangan yang dicari yang kemudian peluang tersebut dibuatkan dalam bentuk table yang sesuai dengan pola juga.

7. Evaluasi Model

   Data yang telah diuji akan di evaluasi dengan menggunakan

   - *Use Testing set* yaitu di dapatkan dari *Data Testing*
   - *Use Training set* percobaan menggunakan *Data Training*
   - 5-Fold Cross Validation yaitu k-fold cross validation dimana percobaan sebanyak k kali untuk satu model dengan parameter yang sama dalam dalam hal ini k yaitu bernilai 5 berarti melakukan percobaan 5 dalam satu model
   - 10-Fold Cross Validation sama seperti 5-fold Cross Validation tetapi percobaan yang dilakukan sebanyak 10 kali dengan menggunakan satu model yang sama

8. *Compered*

   Setelah data di evaluasi kemudian data akan di Compared atau di bandingkan berdasarkan 3 metode yang di uji, dalam kasus ini hal yang akan di uji adalah :

   - Accuracy
   - TPR, FPR,
   - Precision,
   - Recall,
   - Fp-measure

9. *Best Result*

   Metode terbaik berdasarkan pengujian dan evaluasi.

10. *Conclusion*

    Kesimpulan dari semua alur eksperimen yang dilakukan.

3.3 ALAT PENELITIAN

Penulis menggunakan beberapa alat/piranti yang digunakan untuk melakukan pengolahan data/bahan penelitian, yaitu :

1. Hardware, dengan spesifikasi sebagai Berikut:
   a. Laptop, dengan processor core i5
   b. RAM : 8.00 GB
   c. SSD : 500 GB
2. Software, dengan keterangan sebagai berikut:
   a. OS Windows 10 (64 bit)
   b. Microsoft Word & Excel 2019
   c. Weka 3.8.6

3.4 JADWAL PENELITIAN

No	Jenis Kegiatan	Bulan																							
		Sept				Okt				Nov				Des				Jan				Feb			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Prumusan Masalah																								
2	Penentuan Tujuan																								
3	Mempelajari Literatur																								
4	Pengumpulan Data & Informasi																								
5	Proses Data Mining																								
6	Pengujian																								
7	Pembuatan Laporan																								

DAFTAR PUSTAKA

[1] M. A. Ghani and A. Subekti, "Email Spam Filtering Dengan Algoritma Random Forest," *IJCIT (Indonesian J. Comput. Inf. Technol.*, vol. 3, no. 2, pp. 216–221, 2018.

[2] R. Mishra and R. S. Thakur, "Analysis of Random Forest and Naïve Bayes for Spam Mail using Feature Selection Catagorization," *Int. J. Comput. Appl.*, vol. 80, no. 3, pp. 42–47, 2013, doi: 10.5120/13844-1670.

[3] N. F. Rusland, N. Wahid, S. Kasim, and H. Hafit, "Analysis of Naïve Bayes Algorithm for Email Spam Filtering across Multiple Datasets," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 226, no. 1, 2017, doi: 10.1088/1757-899X/226/1/012091.

[4] A. Nur Khormarudin, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Ilmu Komput.*, pp. 1–12, 2016, [Online]. Available: https://ilmukomputer.org/category/datamining/

[5] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.

[6] J. K. Hayes *et al.*, "2222," *Int. J. Radiat. Oncol.*, vol. 66, no. 3, p. S333, 2006, doi: 10.1016/j.ijrobp.2006.07.628.

[7] B. Kurniawan, S. Effendi, and O. S. Sitompul, "Klasifikasi Konten Berita Dengan Metode Text Mining," *J. Dunia Teknol. Inf.*, vol. 1, no. 1, pp. 14–19, 2012, [Online]. Available: http://download.portalgaruda.org/article.php?article=58993&val=4123

[8] B. Laboratories and L. Technologies, "Communication Flows : Analysis of an Email Network Measuring An Email Network," no. July, pp. 1–15, 2014.

[9] A. D. Wibisono, S. Dadi Rizkiono, and A. Wantoro, "Filtering Spam Email

Menggunakan Metode Naive Bayes," *TELEFORTECH J. Telemat. Inf. Technol.*, vol. 1, no. 1, pp. 9–17, 2020, doi: 10.33365/tft.v1i1.685.

[10] N. Y. Septian, "Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro," *J. Semant. 2013*, pp. 1–11, 2009.

[11] V. No, S. Sandra, and A. Heryanto, "Visualisasi Serangan Brute Force Menggunakan Metode K-Means dan Naïve Bayes," vol. 2, no. 1, pp. 315–320, 2016.

[12] A. Harris and A. E. Mintaria, "Komparasi Information Gain , Gain Ratio , CFs-Bestfirst dan CFs-PSO Search Terhadap Performa Deteksi Anomali," vol. 5, pp. 332–343, 2021, doi: 10.30865/mib.v5i1.2258.

[13] S. N. D. Pratiwi and B. S. S. Ulama, "Klasifikasi Email Spam dengan Menggunakan Metode Support Vector Machine dan k-Nearest Neighbor," *J. Sains dan Seni ITS*, vol. 5, no. 2, pp. 344–349, 2016.

[14] A. Z. Farmadiansyah, A. F. Hidayatullah, and F. Rahma, "Deteksi Surel Spam dan Non Spam Bahasa Indonesia," *Automata*, 2021, [Online]. Available: https://journal.uii.ac.id/AUTOMATA/article/view/19514

[15] R. Chandra and D. C. Juan, "Deteksi Pesan Spam pada Forum Daring Menggunakan Metode Naïve Bayes," vol. 7, no. 2, pp. 369–373, 2022.

[16] F. Rozi and R. Kartadie, "Deteksi E-Mail Dan Spam Menggunakan Fuzzy Association Rule Mining," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 2, no. 2, pp. 94–98, 2017, doi: 10.29100/jipi.v2i2.348.

[17] U. Kristen, D. Wacana, and Y. Lukito, "KLASIFIKASI KOMENTAR SPAM PADA INSTAGRAM BERBAHASA INDONESIA MENGGUNAKAN K-NN Thesis View project Sentiment analysis View project Antonius Rachmat," no. February 2018, 2017, [Online]. Available: https://www.researchgate.net/publication/323257693

[18] F. Firdausillah, M. Hafidz, E. D. Udayanti, and E. Kartikadarma, “Sistem Deteksi Surel SPAM Dengan DNSBL Dan Support Vector Machine Pada Penyedia Layanan Mail Marketing,” vol. 3, no. 4, pp. 618–625, 2022, doi: 10.47065/josh.v3i4.1795.

Lampiran Dataset

@attribute text string

@attribute @@class@@ {ham,spam}

@data

'From exmh-workers-admin@redhat.com Thu Aug 22 12:36:23 2002\nReturn-Path: <exmh-workers-admin@example.com>\nDelivered-To: zzzz@localhost.netnoteinc.com\nReceived: from localhost (localhost [127.0.0.1])\n\tby phobos.labs.netnoteinc.com (Postfix) with ESMTP id D03E543C36\n\tfor <zzzz@localhost>; Thu, 22 Aug 2002 07:36:16 -0400 (EDT)\nReceived: from phobos [127.0.0.1]\n\tby localhost with IMAP (fetchmail-5.9.0)\n\tfor zzzz@localhost (single-drop); Thu, 22 Aug 2002 12:36:16 +0100 (IST)\nReceived: from listman.example.com (listman.example.com [66.187.233.211]) by\n dogma.slashnull.org (8.11.6/8.11.6) with ESMTP id g7MBYrZ04811 for\n <zzzz-exmh@example.com>; Thu, 22 Aug 2002 12:34:53 +0100\nReceived: from listman.example.com (localhost.localdomain [127.0.0.1]) by\n listman.redhat.com (Postfix) with ESMTP id 8386540858; Thu, 22 Aug 2002\n 07:35:02 -0400 (EDT)\nDelivered-To: exmh-workers@listman.example.com\nReceived: from int-mx1.corp.example.com (int-mx1.corp.example.com\n [172.16.52.254]) by listman.redhat.com (Postfix) with ESMTP id 10CF8406D7\n for <exmh-workers@listman.redhat.com>; Thu, 22 Aug 2002 07:34:10 -0400\n (EDT)\nReceived: (from mail@localhost) by int-mx1.corp.example.com (8.11.6/8.11.6)\n id g7MBY7g11259 for exmh-workers@listman.redhat.com; Thu, 22 Aug 2002\n 07:34:07 -0400\nReceived: from mx1.example.com (mx1.example.com [172.16.48.31]) by\n int-mx1.corp.redhat.com (8.11.6/8.11.6) with SMTP id g7MBY7Y11255 for\n <exmh-workers@redhat.com>; Thu, 22 Aug 2002 07:34:07 -0400\nReceived: from ratree.psu.ac.th ([202.28.97.6]) by mx1.example.com\n (8.11.6/8.11.6) with

SMTP id g7MBIhl25223 for <exmh-workers@redhat.com>;\n Thu, 22 Aug 2002 07:18:55 -0400\nReceived: from delta.cs.mu.OZ.AU (delta.coe.psu.ac.th [172.30.0.98]) by\n ratree.psu.ac.th (8.11.6/8.11.6) with ESMTP id g7MBWel29762;\n Thu, 22 Aug 2002 18:32:40 +0700 (ICT)\nReceived: from munnari.OZ.AU (localhost [127.0.0.1]) by delta.cs.mu.OZ.AU\n (8.11.6/8.11.6) with ESMTP id g7MBQPW13260; Thu, 22 Aug 2002 18:26:25\n +0700 (ICT)\nFrom: Robert Elz <kre@munnari.OZ.AU>\nTo: Chris Garrigues <cwg-dated-1030377287.06fa6d@DeepEddy.Com>\nCc: exmh-workers@example.com\nSubject: Re: New Sequences Window\nIn-Reply-To: <1029945287.4797.TMDA@deepeddy.vircio.com>\nReferences: <1029945287.4797.TMDA@deepeddy.vircio.com>\n <1029882468.3116.TMDA@deepeddy.vircio.com> <9627.1029933001@munnari.OZ.AU>\n <1029943066.26919.TMDA@deepeddy.vircio.com>\n <1029944441.398.TMDA@deepeddy.vircio.com>\nMIME-Version: 1.0\nContent-Type: text/plain; charset=us-ascii\nMessage-Id: <13258.1030015585@munnari.OZ.AU>\nX-Loop: exmh-workers@example.com\nSender: exmh-workers-admin@example.com\nErrors-To: exmh-workers-admin@example.com\nX-Beenthere: exmh-workers@example.com\nX-Mailman-Version: 2.0.1\nPrecedence: bulk\nList-Help: <mailto:exmh-workers-request@example.com?subject=help>\nList-Post: <mailto:exmh-workers@example.com>\nList-Subscribe: <https://listman.example.com/mailman/listinfo/exmh-workers>,\n <mailto:exmh-workers-request@redhat.com?subject=subscribe>\nList-Id: Discussion list for EXMH developers <exmh-workers.example.com>\nList-Unsubscribe: <https://listman.example.com/mailman/listinfo/exmh-workers>,\n <mailto:exmh-workers-request@redhat.com?subject=unsubscribe>\nList-Archive: <https://listman.example.com/mailman/private/exmh-workers/>\nDate: Thu, 22 Aug 2002 18:26:25 +0700\n\n Date: Wed, 21 Aug 2002 10:54:46 -0500\n From: Chris Garrigues <cwg-dated-1030377287.06fa6d@DeepEddy.Com>\n Message-ID:

<1029945287.4797.TMDA@deepeddy.vircio.com>\n\n\n | I can\'t reproduce this error.\n\nFor me it is very repeatable... (like every time, without fail).\n\nThis is the debug log of the pick happening ...\n\n18:19:03 Pick_It {exec pick +inbox -list -lbrace -lbrace -subject ftp -rbrace -rbrace} {4852-4852 -sequence mercury}\n18:19:03 exec pick +inbox -list -lbrace -lbrace -subject ftp -rbrace -rbrace 4852-4852 -sequence mercury\n18:19:04 Ftoc_PickMsgs {{1 hit}}\n18:19:04 Marking 1 hits\n18:19:04 tkerror: syntax error in expression \"int ...\n\nNote, if I run the pick command by hand ...\n\ndelta$ pick +inbox -list -lbrace -lbrace -subject ftp -rbrace -rbrace 4852-4852 -sequence mercury\n1 hit\n\nThat\'s where the \"1 hit\" comes from (obviously). The version of nmh I\'m\nusing is ...\n\ndelta$ pick -version\npick -- nmh-1.0.4 [compiled on fuchsia.cs.mu.OZ.AU at Sun Mar 17 14:55:56 ICT 2002]\n\nAnd the relevant part of my .mh_profile ...\n\ndelta$ mhparam pick\n-seq sel -list\n\n\nSince the pick command works, the sequence (actually, both of them, the\none that\'s explicit on the command line, from the search popup, and the\none that comes from .mh_profile) do get created.\n\nkre\n\nps: this is still using the version of the code form a day ago, I haven\'t\nbeen able to reach the cvs repository today (local routing issue I think).\n\n\n\n_____\nExmh-workers mailing list\nExmh-workers@redhat.com\nhttps://listman.redhat.com/mailman/listinfo/exmh-workers\n\n',ham

'From Steve_Burt@cursor-system.com Thu Aug 22 12:46:39 2002\nReturn-Path: <Steve_Burt@cursor-system.com>\nDelivered-To: zzzz@localhost.netnoteinc.com\nReceived: from localhost (localhost [127.0.0.1])\n\tby phobos.labs.netnoteinc.com (Postfix) with ESMTP id BE12E43C34\n\tfor <zzzz@localhost>; Thu, 22 Aug 2002 07:46:38 -0400 (EDT)\nReceived: from phobos [127.0.0.1]\n\tby localhost with IMAP (fetchmail-5.9.0)\n\tfor zzzz@localhost (single-drop); Thu, 22 Aug 2002 12:46:38 +0100 (IST)\nReceived: from n20.grp.scd.yahoo.com (n20.grp.scd.yahoo.com\n [66.218.66.76]) by dogma.slashnull.org (8.11.6/8.11.6) with SMTP id\n

g7MBkTZ05087 for <zzzz@example.com>; Thu, 22 Aug 2002 12:46:29 +0100\nX-Egroups-Return: sentto-2242572-52726-1030016790-zzzz=example.com@returns.groups.yahoo.com\nReceived: from [66.218.67.196] by n20.grp.scd.yahoo.com with NNFMP;\n 22 Aug 2002 11:46:30 -0000\nX-Sender: steve.burt@cursor-system.com\nX-Apparently-To: zzzzteana@yahoogroups.com\nReceived: (EGP: mail-8_1_0_1); 22 Aug 2002 11:46:29 -0000\nReceived: (qmail 11764 invoked from network); 22 Aug 2002 11:46:29 -0000\nReceived: from unknown (66.218.66.217) by m3.grp.scd.yahoo.com with QMQP;\n 22 Aug 2002 11:46:29 -0000\nReceived: from unknown (HELO mailgateway.cursor-system.com) (62.189.7.27)\n by mta2.grp.scd.yahoo.com with SMTP; 22 Aug 2002 11:46:29 -0000\nReceived: from exchange1.cps.local (unverified) by\n mailgateway.cursor-system.com (Content Technologies SMTPRS 4.2.10) with\n ESMTP id <T5cde81f695ac1d100407d@mailgateway.cursor-system.com> for\n <forteana@yahoogroups.com>; Thu, 22 Aug 2002 13:14:10 +0100\nReceived: by exchange1.cps.local with Internet Mail Service (5.5.2653.19)\n id <PXX6AT23>; Thu, 22 Aug 2002 12:46:27 +0100\nMessage-Id: <5EC2AD6D2314D14FB64BDA287D25D9EF12B4F6@exchange1.cps.local>\n To: \"\'zzzzteana@yahoogroups.com\'\" <zzzzteana@yahoogroups.com>\nX-Mailer: Internet Mail Service (5.5.2653.19)\nX-Egroups-From: Steve Burt <steve.burt@cursor-system.com>\nFrom: Steve Burt <Steve_Burt@cursor-system.com>\nX-Yahoo-Profile: pyruse\nMIME-Version: 1.0\nMailing-List: list zzzzteana@yahoogroups.com; contact\n forteana-owner@yahoogroups.com\nDelivered-To: mailing list zzzzteana@yahoogroups.com\nPrecedence: bulk\nList-Unsubscribe: <mailto:zzzzteana-unsubscribe@yahoogroups.com>\nDate: Thu, 22 Aug 2002 12:46:18 +0100\nSubject: [zzzzteana] RE: Alexander\nReply-To: zzzzteana@yahoogroups.com\nContent-Type: text/plain; charset=US-ASCII\nContent-Transfer-Encoding: 7bit\n\nMartin A posted:\nTassos Papadopoulos, the Greek sculptor behind the plan, judged that the\n limestone of Mount Kerdylio, 70 miles east of Salonika and not far from the\n Mount Athos

monastic community, was ideal for the patriotic sculpture. \n \n As well as Alexander\'s granite features, 240 ft high and 170 ft wide, a\n museum, a restored amphitheatre and car park for admiring crowds are\nplanned\n----------\nSo is this mountain limestone or granite?\nIf it\'s limestone, it\'ll weather pretty fast.\n\n---------- Yahoo! Groups Sponsor ----------~-->\n4 DVDs Free +s&p Join Now\nhttp://us.click.yahoo.com/pt6YBB/NXiEAA/mG3HAA/7gSolB/TM\n----------~->\n\nTo unsubscribe from this group, send an email to:\nforteana-unsubscribe@egroups.com\n\n \n\nYour use of Yahoo! Groups is subject to http://docs.yahoo.com/info/terms/ \n\n\n\n',ham

'From timc@2ubh.com Thu Aug 22 13:52:59 2002\nReturn-Path: <timc@2ubh.com>\nDelivered-To: zzzz@localhost.netnoteinc.com\nReceived: from localhost (localhost [127.0.0.1])\n\tby phobos.labs.netnoteinc.com (Postfix) with ESMTP id 0314547C66\n\tfor <zzzz@localhost>; Thu, 22 Aug 2002 08:52:58 -0400 (EDT)\nReceived: from phobos [127.0.0.1]\n\tby localhost with IMAP (fetchmail-5.9.0)\n\tfor zzzz@localhost (single-drop); Thu, 22 Aug 2002 13:52:59 +0100 (IST)\nReceived: from n16.grp.scd.yahoo.com (n16.grp.scd.yahoo.com\n [66.218.66.71]) by dogma.slashnull.org (8.11.6/8.11.6) with SMTP id\n g7MCrdZ07070 for <zzzz@example.com>; Thu, 22 Aug 2002 13:53:39 +0100\nX-Egroups-Return: sentto-2242572-52733-1030020820-zzzz=example.com@returns.groups.yahoo.com\nReceived: from [66.218.67.198] by n16.grp.scd.yahoo.com with NNFMP;\n 22 Aug 2002 12:53:40 -0000\nX-Sender: timc@2ubh.com\nX-Apparently-To: zzzzteana@yahoogroups.com\nReceived: (EGP: mail-8_1_0_1); 22 Aug 2002 12:53:39 -0000\nReceived: (qmail 76099 invoked from network); 22 Aug 2002 12:53:39 -0000\nReceived: from unknown (66.218.66.218) by m5.grp.scd.yahoo.com with QMQP;\n 22 Aug 2002 12:53:39 -0000\nReceived: from unknown (HELO rhenium.btinternet.com) (194.73.73.93) by\n mta3.grp.scd.yahoo.com with SMTP; 22 Aug 2002 12:53:39 -0000\nReceived: from host217-36-23-185.in-addr.btopenworld.com ([217.36.23.185])\n by

rhenium.btinternet.com with esmtp (Exim 3.22 #8) id 17hrT0-0004gj-00\n for forteana@yahoogroups.com; Thu, 22 Aug 2002 13:53:38 +0100\nX-Mailer: Microsoft Outlook Express Macintosh Edition - 4.5 (0410)\nTo: zzzzteana <zzzzteana@yahoogroups.com>\nX-Priority: 3\nMessage-Id: <E17hrT0-0004gj-00@rhenium.btinternet.com>\nFrom: \"Tim Chapman\" <timc@2ubh.com>\nX-Yahoo-Profile: tim2ubh\nMIME-Version: 1.0\nMailing-List: list zzzzteana@yahoogroups.com; contact\n forteana-owner@yahoogroups.com\nDelivered-To: mailing list zzzzteana@yahoogroups.com\nPrecedence: bulk\nList-Unsubscribe: <mailto:zzzzteana-unsubscribe@yahoogroups.com>\nDate: Thu, 22 Aug 2002 13:52:38 +0100\nSubject: [zzzzteana] Moscow bomber\nReply-To: zzzzteana@yahoogroups.com\nContent-Type: text/plain; charset=US-ASCII\nContent-Transfer-Encoding: 7bit\n\nMan Threatens Explosion In Moscow \n\nThursday August 22, 2002 1:40 PM\nMOSCOW (AP) - Security officers on Thursday seized an unidentified man who\nsaid he was armed with explosives and threatened to blow up his truck in\nfront of Russia\'s Federal Security Services headquarters in Moscow, NTV\ntelevision reported.\nThe officers seized an automatic rifle the man was carrying, then the man\ngot out of the truck and was taken into custody, NTV said. No other details\nwere immediately available.\nThe man had demanded talks with high government officials, the Interfax and\nITAR-Tass news agencies said. Ekho Moskvy radio reported that he wanted to\ntalk with Russian President Vladimir Putin.\nPolice and security forces rushed to the Security Service building, within\nblocks of the Kremlin, Red Square and the Bolshoi Ballet, and surrounded the\nman, who claimed to have one and a half tons of explosives, the news\nagencies said. Negotiations continued for about one and a half hours outside\nthe building, ITAR-Tass and Interfax reported, citing witnesses.\nThe man later drove away from the building, under police escort, and drove\nto a street near Moscow\'s Olympic Penta Hotel, where authorities held\nfurther negotiations with him, the Moscow police press service said. The\nmove appeared to be an attempt by security services to get him to a more\nsecure location. \n\n---------- Yahoo! Groups Sponsor ----------~-->\n4 DVDs Free +s&p Join

Now\nhttp://us.click.yahoo.com/pt6YBB/NXiEAA/mG3HAA/7gSolB/TM\n---~->\n\nTo unsubscribe from this group, send an email to:\nforteana-unsubscribe@egroups.com\n\n \n\nYour use of Yahoo! Groups is subject to http://docs.yahoo.com/info/terms/ \n\n\n\n',ham

'From irregulars-admin@tb.tf Thu Aug 22 14:23:39 2002\nReturn-Path: <irregulars-admin@tb.tf>\nDelivered-To: zzzz@localhost.netnoteinc.com\nReceived: from localhost (localhost [127.0.0.1])\n\tby phobos.labs.netnoteinc.com (Postfix) with ESMTP id 9DAE147C66\n\tfor <zzzz@localhost>; Thu, 22 Aug 2002 09:23:38 -0400 (EDT)\nReceived: from phobos [127.0.0.1]\n\tby localhost with IMAP (fetchmail-5.9.0)\n\tfor zzzz@localhost (single-drop); Thu, 22 Aug 2002 14:23:38 +0100 (IST)\nReceived: from web.tb.tf (route-64-131-126-36.telocity.com\n [64.131.126.36]) by dogma.slashnull.org (8.11.6/8.11.6) with ESMTP id\n g7MDGOZ07922 for <zzzz-irr@example.com>; Thu, 22 Aug 2002 14:16:24 +0100\nReceived: from web.tb.tf (localhost.localdomain [127.0.0.1]) by web.tb.tf\n (8.11.6/8.11.6) with ESMTP id g7MDP9I16418; Thu, 22 Aug 2002 09:25:09\n -0400\nReceived: from red.harvee.home (red [192.168.25.1] (may be forged)) by\n web.tb.tf (8.11.6/8.11.6) with ESMTP id g7MDO4I16408 for\n <irregulars@tb.tf>; Thu, 22 Aug 2002 09:24:04 -0400\nReceived: from prserv.net (out4.prserv.net [32.97.166.34]) by\n red.harvee.home (8.11.6/8.11.6) with ESMTP id g7MDFBD29237 for\n <irregulars@tb.tf>; Thu, 22 Aug 2002 09:15:12 -0400\nReceived: from [209.202.248.109]\n (slip-32-103-249-10.ma.us.prserv.net[32.103.249.10]) by prserv.net (out4)\n with ESMTP id <20020822131502204 05qu8jce>; Thu, 22 Aug 2002 13:15:07 +0000\nMIME-Version: 1.0\nX-Sender: @ (Unverified)\nMessage-Id: <p04330137b98a941c58a8@[209.202.248.109]>\nTo: undisclosed-recipient: ;\nFrom: Monty Solomon <monty@roscom.com>\nContent-Type: text/plain; charset=\"us-ascii\"\nSubject: [IRR] Klez: The Virus That Won\'t Die\nSender: irregulars-admin@tb.tf\nErrors-To: irregulars-admin@tb.tf\nX-Beenthere: irregulars@tb.tf\nX-Mailman-Version: 2.0.6\nPrecedence: bulk\nList-Help:

<mailto:irregulars-request@tb.tf?subject=help>\nList-Post: <mailto:irregulars@tb.tf>\nList-Subscribe: <http://tb.tf/mailman/listinfo/irregulars>,\n <mailto:irregulars-request@tb.tf?subject=subscribe>\nList-Id: New home of the TBTF Irregulars mailing list <irregulars.tb.tf>\nList-Unsubscribe: <http://tb.tf/mailman/listinfo/irregulars>,\n <mailto:irregulars-request@tb.tf?subject=unsubscribe>\nList-Archive: <http://tb.tf/mailman/private/irregulars/>\nDate: Thu, 22 Aug 2002 09:15:25 -0400\n\nKlez: The Virus That Won\'t Die\n \nAlready the most prolific virus ever, Klez continues to wreak havoc.\n\nAndrew Brandt\n>>From the September 2002 issue of PC World magazine\nPosted Thursday, August 01, 2002\n\n\nThe Klez worm is approaching its seventh month of wriggling across \nthe Web, making it one of the most persistent viruses ever. And \nexperts warn that it may be a harbinger of new viruses that use a \ncombination of pernicious approaches to go from PC to PC.\n\nAntivirus software makers Symantec and McAfee both report more than \n2000 new infections daily, with no sign of letup at press time. The \nBritish security firm MessageLabs estimates that 1 in every 300 \ne-mail messages holds a variation of the Klez virus, and says that \nKlez has already surpassed last summer\'s SirCam as the most prolific \nvirus ever.\n\nAnd some newer Klez variants aren\'t merely nuisances--they can carry \nother viruses in them that corrupt your data.\n\n...\n\nhttp://www.pcworld.com/news/article/0,aid,103259,00.asp\n_____ _____\nIrregulars mailing list\nIrregulars@tb.tf\nhttp://tb.tf/mailman/listinfo/irregulars\n\n',ham

'From exmh-users-admin@redhat.com Thu Aug 22 14:44:07 2002\nReturn-Path: <exmh-users-admin@example.com>\nDelivered-To: zzzz@localhost.netnoteinc.com\nReceived: from localhost (localhost [127.0.0.1])\n\tby phobos.labs.netnoteinc.com (Postfix) with ESMTP id 985B247C67\n\tfor <zzzz@localhost>; Thu, 22 Aug 2002 09:44:04 -0400 (EDT)\nReceived: from phobos [127.0.0.1]\n\tby localhost with IMAP (fetchmail-

5.9.0)\n\tfor zzzz@localhost (single-drop); Thu, 22 Aug 2002 14:44:04 +0100 (IST)\nReceived: from listman.example.com (listman.example.com [66.187.233.211]) by\n dogma.slashnull.org (8.11.6/8.11.6) with ESMTP id g7MDgEZ08598 for\n <zzzz-exmh@example.com>; Thu, 22 Aug 2002 14:42:19 +0100\nReceived: from listman.example.com (localhost.localdomain [127.0.0.1]) by\n listman.redhat.com (Postfix) with ESMTP id F26113EE9A; Thu, 22 Aug 2002\n 09:42:15 -0400 (EDT)\nDelivered-To: exmh-users@listman.example.com\nReceived: from int-mx1.corp.example.com (int-mx1.corp.example.com\n [172.16.52.254]) by listman.redhat.com (Postfix) with ESMTP id 4ACEF3F4A2\n for <exmh-users@listman.redhat.com>; Thu, 22 Aug 2002 09:38:03 -0400 (EDT)\nReceived: (from mail@localhost) by int-mx1.corp.example.com (8.11.6/8.11.6)\n id g7MDc0601591 for exmh-users@listman.redhat.com; Thu, 22 Aug 2002\n 09:38:00 -0400\nReceived: from mx1.example.com (mx1.example.com [172.16.48.31]) by\n int-mx1.corp.redhat.com (8.11.6/8.11.6) with SMTP id g7MDc0Y01587 for\n <exmh-users@redhat.com>; Thu, 22 Aug 2002 09:38:00 -0400\nReceived: from mta03bw.bigpond.com (mta03bw.bigpond.com [139.134.6.86]) by\n mx1.redhat.com (8.11.6/8.11.6) with SMTP id g7MDNVl14108 for\n <exmh-users@redhat.com>; Thu, 22 Aug 2002 09:23:31 -0400\nReceived: from hobbit.linuxworks.com.au ([144.135.24.81]) by\n mta03bw.bigpond.com (Netscape Messaging Server 4.15 mta03bw May 23 2002\n 23:53:28) with SMTP id H18Z7300.F6G for <exmh-users@redhat.com>;\n Thu, 22 Aug 2002 23:37:51 +1000\nReceived: from CPE-203-51-220-31.qld.bigpond.net.au ([203.51.220.31]) by\n bwmam05.mailsvc.email.bigpond.com(MailRouter V3.0n 44/32989362);\n 22 Aug 2002 23:37:51\nReceived: (from tony@localhost) by hobbit.linuxworks.com.au\n (8.11.6/8.11.6) id g7MDaWX26868; Thu, 22 Aug 2002 23:36:32 +1000\nMessage-Id: <200208221336.g7MDaWX26868@hobbit.linuxworks.com.au.nospam>\nTo: Exmh Users Mailing List <exmh-users@example.com>\nFrom: Tony Nugent <tony@linuxworks.com.au>\nX-Face:]IrGs{LrofDtGfsrG!As5=G\'2HRr2zt:H>djXb5@v|Dr!jOelxzAZ`!}(\"]}]\n

Q!)1w#X;)nLlb\'XhSu,QL>;)L/l06wsI?rv-xy6\%Y1e\"BUiV\%)mU;]f-5<#U6\n UthZ0QrF7_p#q}*Cn}jd|XT~7P7ik]Q!2u\%aTtvc;)zfH\\:3f<[a:)M\nOrganizatio n: Linux Works for network\nX-Mailer: nmh-1.0.4 exmh-2.4\nX-Os: Linux-2.4 RedHat 7.2\nIn-Reply-To: message-id <200208212046.g7LKkqf15798@mail.banirh.com> of Wed,\n Aug 21 15:46:52 2002\nSubject: Re: Insert signature\nX-Loop: exmh-users@example.com\nSender: exmh-users-admin@example.com\nErrors-To: exmh-users-admin@example.com\nX-Beenthere: exmh-users@example.com\nX-Mailman-Version: 2.0.1\nPrecedence: bulk\nReply-To: exmh-users@example.com\nList-Help: <mailto:exmh-users-request@example.com?subject=help>\nList-Post: <mailto:exmh-users@example.com>\nList-Subscribe: <https://listman.example.com/mailman/listinfo/exmh-users>,\n <mailto:exmh-users-request@redhat.com?subject=subscribe>\nList-Id: Discussion list for EXMH users <exmh-users.example.com>\nList-Unsubscribe: <https://listman.example.com/mailman/listinfo/exmh-users>,\n <mailto:exmh-users-request@redhat.com?subject=unsubscribe>\nList-Archive: <https://listman.example.com/mailman/private/exmh-users/>\nDate: Thu, 22 Aug 2002 23:36:32 +1000\n\nOn Wed Aug 21 2002 at 15:46, Ulises Ponce wrote:\n\n> Hi!\n> \n> Is there a command to insert the signature using a combination of keys and not\n> to have sent the mail to insert it then?\n\nI simply put it (them) into my (nmh) component files (components,\nreplcomps, forwcomps and so on). That way you get them when you are\nediting your message. Also, by using comps files for specific\nfolders you can alter your .sig per folder (and other tricks). See\nthe docs for (n)mh for all the details.\n\nThere might (must?) also be a way to get sedit to do it, but I\'ve\nbeen using gvim as my exmh message editor for a long time now. I\nload it with a command that loads some email-specific settings, eg,\nto \"syntax\" colour-highlight the headers and quoted parts of an\nemail)... it would be possible to map some (vim) keys that would add\na sig (or even give a selection of sigs to choose from).\n\nAnd there are all sorts of ways to have randomly-chosen sigs...\nsomewhere at rtfm.mit.edu... ok, here we go:\nrtfm.mit.edu/pub/usenet-by-

group/news.answers/signature_finger_faq.\n(Warning... it\'s old, May 1995).\n\n> Regards,\n> Ulises\n\nHope this helps.\n\nCheers\nTony\n\n\n\n_____\nExmh-users mailing list\nExmh-users@redhat.com\nhttps://listman.redhat.com/mailman/listinfo/exmh-users\n\n',ham

'From Stewart.Smith@ee.ed.ac.uk Thu Aug 22 14:44:26 2002\nReturn-Path: <Stewart.Smith@ee.ed.ac.uk>\nDelivered-To: zzzz@localhost.netnoteinc.com\nReceived: from localhost (localhost [127.0.0.1])\n\tby phobos.labs.netnoteinc.com (Postfix) with ESMTP id EC69D47C66\n\tfor <zzzz@localhost>; Thu, 22 Aug 2002 09:44:25 -0400 (EDT)\nReceived: from phobos [127.0.0.1]\n\tby localhost with IMAP (fetchmail-5.9.0)\n\tfor zzzz@localhost (single-drop); Thu, 22 Aug 2002 14:44:25 +0100 (IST)\nReceived: from n6.grp.scd.yahoo.com (n6.grp.scd.yahoo.com [66.218.66.90])\n by dogma.slashnull.org (8.11.6/8.11.6) with SMTP id g7MDcOZ08504 for\n <zzzz@example.com>; Thu, 22 Aug 2002 14:38:25 +0100\nX-Egroups-Return: sentto-2242572-52736-1030023506-zzzz=example.com@returns.groups.yahoo.com\nReceived: from [66.218.67.192] by n6.grp.scd.yahoo.com with NNFMP;\n 22 Aug 2002 13:38:26 -0000\nX-Sender: Stewart.Smith@ee.ed.ac.uk\nX-Apparently-To: zzzzteana@yahoogroups.com\nReceived: (EGP: mail-8_1_0_1); 22 Aug 2002 13:38:25 -0000\nReceived: (qmail 48882 invoked from network); 22 Aug 2002 13:38:25 -0000\nReceived: from unknown (66.218.66.218) by m10.grp.scd.yahoo.com with QMQP;\n 22 Aug 2002 13:38:25 -0000\nReceived: from unknown (HELO postbox.ee.ed.ac.uk) (129.215.80.253) by\n mta3.grp.scd.yahoo.com with SMTP; 22 Aug 2002 13:38:24 -0000\nReceived: from ee.ed.ac.uk (sxs@dunblane [129.215.34.86]) by\n postbox.ee.ed.ac.uk (8.11.0/8.11.0) with ESMTP id g7MDcNi28645 for\n <forteana@yahoogroups.com>; Thu, 22 Aug 2002 14:38:23 +0100 (BST)\nMessage-Id: <3D64E94E.8060301@ee.ed.ac.uk>\nOrganization: Scottish

Microelectronics Centre\nUser-Agent: Mozilla/5.0 (X11; U; SunOS sun4u; en-US; rv:1.1b) Gecko/20020628\nX-Accept-Language: en, en-us\nTo: zzzzteana@yahoogroups.com\nReferences: <3D64F325.11319.61EA648@localhost>\nFrom: Stewart Smith <Stewart.Smith@ee.ed.ac.uk>\nX-Yahoo-Profile: stochasticus\nMIME-Version: 1.0\nMailing-List: list zzzzteana@yahoogroups.com; contact\n forteana-owner@yahoogroups.com\nDelivered-To: mailing list zzzztean

Sumber Dataset :

https://spamassassin.apache.org/old/publiccorpus/