

# **IMPLEMENTASI METODE NAIVE BAYES DALAM PENDETEKSIAN EMAIL SPAM & NON-SPAM**

*Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Strata-1 Pada  
Jurusan Teknik Informatika Bilingual Fakultas Ilmu Komputer*



Oleh:

Winarto Willyam

NIM : 09111402016

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2018**

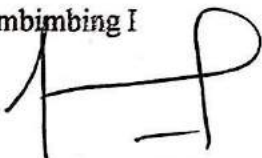
# LEMBAR PENGESAHAN TUGAS AKHIR

## IMPLEMENTASI METODE NAIVE BAYES DALAM PENDETEKSIAN EMAIL SPAM & NON-SPAM

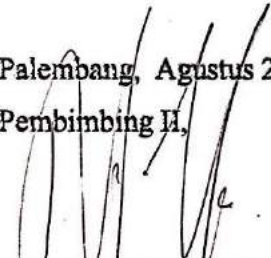
Oleh :

Winarto Willyam  
NIM : 09111402016

Pembimbing I

  
M. Fachrurrozi, M.T.  
NIP 198005222008121002

Palembang, Agustus 2018  
Pembimbing II,

  
Novi Yusliani, M.T.  
NIP 198211082012122001

Mengetahui,  
Ketua Jurusan Teknik Informatika

  
Rifkie Primartha, M.T  
NIP 197706012009121004

## TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Selasa tanggal 31 Juli 2018 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

**N a m a** : Winarto Willyam  
**N I M** : 09111402016  
**J u d u l** : Implementasi Metode Naïve Bayes dalam Pendeteksian Email Spam & Non-Spam

1. Ketua Penguji

M. Fachrurrozi M.T.  
NIP. 198005222008121002

2. Penguji I

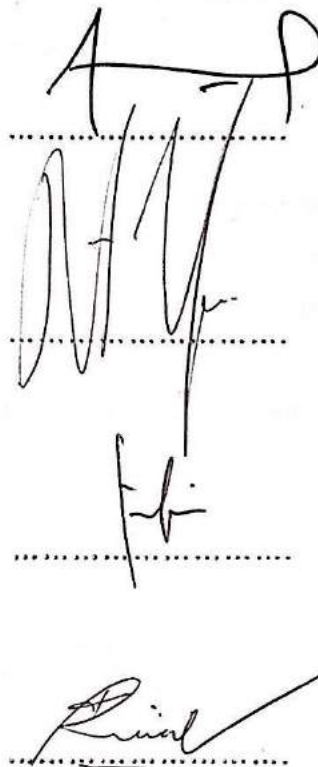
Novi Yusliani, M.T.  
NIP. 198211082012122001

3. Penguji II

Rifkie Primartha, M.T.  
NIP.197706012009121004


4. Penguji III

Mastura Diana Marieska, M.T  
NIP. 198603212018032001



The image shows four handwritten signatures, each on a dotted line. The first signature is for M. Fachrurrozi M.T., the second for Novi Yusliani, M.T., the third for Rifkie Primartha, M.T., and the fourth for Mastura Diana Marieska, M.T.

Mengetahui,  
Ketua Jurusan Teknik Informatika

  
Rifkie Primartha, M. T.  
NIP. 197706012009121004

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Winarto Willyam  
NIM : 09111402016  
Program Studi : Teknik Informatika (Bilingual)  
Judul Skripsi : Implementasi Metode *Naïve Bayes* dalam Pendeteksian *Email Spam & Non-Spam*

Hasil Pengecekan Software *iThenticate/Turnitin*: 18 %

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, Agustus 2018



Winarto Willyam  
NIM. 09111402016

## Motto :

- *“Your time is limited, so dont waste it living someone else's life” – Steve Jobs*
- *“Ada dua hal yang tidak bisa ditunda dalam kehidupan ini, yaitu berbakti kepada orang tua dan melakukan kebajikan.” – Biksu Shih Cheng Yen*

*Kupersembahkan karya tulis ini kepada :*

- *Kedua orang tuaku Bp. William Pratama dan Ibu Susanty Lim*
- *Saudara Perempuan dan Saudara laki-lakiku*
- *Kedua pembimbing dan Dosenku*
- *Semua guru dan karyawan SMA Xaverius 3*
- *Teman-teman IF Bilingual 2011, dan*
- *Keluarga Mahasiswa Buddhis Palembang*

# AN IMPLEMENTATION OF NAÏVE BAYES METHOD ON EMAIL SPAM AND NON-SPAM DETECTION

By :

Winarto Willyam

NIM : 09111402016

## **ABSTRACT**

The spam detection in email is a system that separates or groups messages in electronic mail. Nowadays research and application of this system is growing and becoming more popular. Email is still a powerful tool aimed at the exchange of ideas and information, the cheapest form of communication and able to reach millions of users with a short time. Each email has different message content from private to commercial messages. In this research, there are several processes performed to detect spam that is pre-processing with tokenizing & case folding, feature extraction using Chi Square method and conduct training and testing with Naive Bayes method. This study aims to measure and prove accuracy in detecting spam and non-spam emails using the Naive Bayes method.

Keywords : Spam Detection, Machine Learning, Tokenizing, Case Folding, Email, Bayesian Classifier, Naive Bayes.

# IMPLEMENTASI METODE NAIVE BAYES DALAM PENDETEKSIAN EMAIL SPAM & NON-SPAM

Oleh :

Winarto Willyam

NIM : 09111402016

## ABSTRAK

Sistem pendeteksi spam pada email adalah sistem yang memisahkan atau mengelompokkan pesan dalam surel elektronik. Sekarang penelitian dan penerapan sistem ini menjadi populer dan semakin berkembang. Email masih menjadi alat yang kuat ditujukan untuk pertukaran ide dan informasi, bentuk komunikasi termurah dan mampu menjangkau jutaan pengguna dengan waktu yang singkat. Setiap email memiliki isi pesan yang berbeda-beda dari pesan pribadi hingga komersial. Dalam penelitian ini, ada beberapa proses yang dilakukan untuk mendeteksi spam yaitu pre-processing dengan tokenizing & case folding, ekstraksi fitur menggunakan metode Chi Square dan melakukan pelatihan dan testing dengan metode *Naive Bayes*. Penelitian ini bertujuan untuk mengukur dan membuktikan akurasi dalam mendeteksi email spam dan non-spam dengan metode *Naive Bayes*.

Kata kunci : Deteksi *Spam*, *Machine Learning*, *Tokenizing*, *Case Folding*, *E-mail*, *Bayesian Classifier*, *Naive Bayes*.

## KATA PENGANTAR

Puji syukur bagi Sang Triratna, Para Buddha dan Bodhisattva, atas berkah dan anugerah-Nya sehingga penulis dapat menyelesaikan tugas akhir ini. Tugas akhir yang berjudul **“Implementasi Metode Naïve Bayes dalam Pendeteksian Email Spam & Non-Spam”** ini disusun untuk memenuhi salah satu persyaratan kelulusan tingkat S1 pada Jurusan Teknik Informatika Universitas Sriwijaya.

Pada kesempatan ini, penulis ingin menyampaikan ucapan terimakasih yang tak terhingga kepada pihak-pihak yang telah memberikan dukungan, bimbingan dan motivasi kepada penulis untuk menyelesaikan tugas akhir ini, yaitu kepada:

1. Mamaku tersayang Susanty Lim, Papaku Willyam Pratama, Saudara Perempuan Ellia Santy, Adik laki-lakiku Yulius William dan seluruh keluarga besar atas semua bantuan dan doanya untuk penulis selama ini;
2. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya;
3. Bapak M. Fachrurrozi, M.T. dan Ibu Novi Yusliani, M.T. selaku pembimbing yang telah membimbing dan membantu penulis dengan sabar;
4. Bapak Rifkie Primartha, M.T. dan Ibu Mastura Diana Marieska, M.T. selaku penguji;
5. Ibu Novi Yusliani, M.T. selaku pembimbing akademik penulis, yang sangat baik hati;
6. Bapak dan Ibu dosen yang telah membagikan ilmunya kepada penulis selama belajar di Fakultas Ilmu Komputer Universitas Sriwijaya;



7. Devinta Fulvia Alvianji yang selalu setia memberikan motivasi kepada penulis dalam menyelesaikan tugas akhir;
8. Muhammad Fajriandi, Muhammad Zaki Tamimy, Annisa Ayu Pratiwi, Dyah Lindung Pengasih dan Intan Widyaningrum yang setia menjadi tempat penulis bertukar pendapat selama proses mengerjakan tugas akhir;
9. Teman-teman dari Keluarga Mahasiswa Buddhis Palembang (KMBP) yang selalu memberikan support dan bantuan selama proses mengerjakan tugas akhir;
10. Guruku yang kusayangi, khususnya Bapak Andreas Toni Hartono, S.Pd. yang selalu memberikan support selama proses mengerjakan tugas akhir.
11. Teman-teman IFBIL 2011;
12. Untuk semua pihak yang telah membantu penyelesaian tugas akhir ini dan tidak dapat disebutkan satu persatu.

Akhir kata, penulis menyadari bahwa tugas akhir ini jauh dari kata sempurna. Untuk itu penulis mengharapkan kritik dan saran yang membangun dari semua pihak untuk penyempurnaan tugas akhir ini. Dan semoga tugas akhir ini dapat bermanfaat bagi pihak yang membutuhkan.

Sadhu... Sadhu... Sadhu...

Palembang, Agustus 2018

Penulis

## DAFTAR ISI

Halaman

|   |       |
|---|-------|
| <b>HALAMAN JUDUL</b> .....                          | i     |
| <b>HALAMAN PENGESAHAN</b> .....                     | ii    |
| <b>HALAMAN TANDA LULUS UJIAN SIDANG AKHIR</b> ..... | iii   |
| <b>HALAMAN PERNYATAAN BEBAS PLAGIAT</b> .....       | iv    |
| <b>MOTTO DAN PERSEMBAHAN</b> .....                  | v     |
| <b>ABSTRACT</b> .....                               | vi    |
| <b>ABSTRAK</b> .....                                | vii   |
| <b>KATA PENGANTAR</b> .....                         | viii  |
| <b>DAFTAR ISI</b> .....                             | x     |
| <b>DAFTAR GAMBAR</b> .....                          | xiv   |
| <b>DAFTAR TABEL</b> .....                           | xvi   |
| <b>DAFTAR LAMPIRAN</b> .....                        | xviii |

### **BAB I PENDAHULUAN**

|                                  |     |
|----------------------------------|-----|
| 1.1. Pendahuluan .....           | I-1 |
| 1.2. Latar Belakang Masalah..... | I-1 |
| 1.3. Rumusan Masalah .....       | I-3 |
| 1.4. Tujuan Penelitian.....      | I-4 |
| 1.5. Manfaat Penelitian.....     | I-4 |
| 1.6. Batasan Masalah.....        | I-5 |
| 1.7. Sistematika Penulisan.....  | I-5 |
| 1.8. Kesimpulan.....             | I-7 |

### **BAB II KAJIAN LITERATUR**

|                                     |      |
|-------------------------------------|------|
| 2.1. Pendahuluan .....              | II-1 |
| 2.2. Landasan Teori.....            | II-1 |
| 2.2.1. <i>Electronic Mail</i> ..... | II-1 |
| 2.2.1.1. <i>Spam</i> .....          | II-2 |

|  |       |
|--|-------|
| 2.2.1.2. <i>Non-Spam</i> .....                     | II-4  |
| 2.2.2. <i>Pre-Processing</i> .....                 | II-4  |
| 2.2.3. <i>Bayesian Classifier</i> .....            | II-5  |
| 2.2.4. <i>Rational Unified Process (RUP)</i> ..... | II-7  |
| 2.3. Penelitian Lain yang Relevan.....             | II-12 |
| 2.4. Kesimpulan.....                               | II-13 |

### **BAB III METODOLOGI PENELITIAN**

|  |        |
|--|--------|
| 3.1. Pendahuluan .....   | III-1  |
| 3.2. Metode Pengumpulan Data .....                                     | III-1  |
| 3.2.1. Jenis Data .....  | III-1  |
| 3.2.2. Sumber Data .....   | III-1  |
| 3.2.3. Metode Pengumpulan Data .....                                   | III-2  |
| 3.3. Tahapan Penelitian .....  | III-2  |
| 3.3.1. Pra Pemrosesan dengan <i>Tokenizing</i> dan <i>Case Folding</i> | III-2  |
| 3.3.2. Pengenalan Spam dengan Metode Naïve Bayes .                     | III-5  |
| 3.3.3. Pengujian Penelitian .....                                      | III-15 |
| 3.4. Metode Pengembangan Perangkat Lunak .....                         | III-19 |
| 3.4.1. Fase Insepsi .....  | III-20 |
| 3.4.2. Fase Elaborasi.....   | III-20 |
| 3.4.3. Fase Konstruksi .....   | III-20 |
| 3.4.4. Fase Transisi .....   | III-21 |
| 3.5. Penjadwalan Penelitian .....                                      | III-21 |
| 3.6. Kesimpulan.....   | III-30 |

### **BAB IV PENGEMBANGAN PERANGKAT LUNAK**

|  |      |
|--|------|
| 4.1. Pendahuluan .....                 | IV-1 |
| 4.2. Fase Insepsi .....                | IV-1 |
| 4.2.1. Kebutuhan Perangkat Lunak ..... | IV-1 |

|          |   |       |
|----------|---|-------|
| 4.2.2.   | Pemodelan <i>Use Case</i> .....   | IV-2  |
| 4.2.2.1. | Diagram <i>Use Case</i> .....   | IV-3  |
| 4.2.2.2. | Definisi Pengguna .....   | IV-3  |
| 4.2.2.3. | Definisi <i>Use Case</i> .....  | IV-4  |
| 4.2.2.4. | Skenario <i>Use Case</i> .....  | IV-4  |
| 4.2.2.5. | Kelas Analisis .....  | IV-5  |
| 4.3.     | Fase Elaborasi .....  | IV-7  |
| 4.3.1.   | Perancangan Antarmuka.....  | IV-7  |
| 4.3.2.   | Diagram Sekuensial.....   | IV-8  |
| 4.3.2.1. | Diagram Sequence Mendeteksi <i>Email</i><br>dengan <i>Naïve Bayes</i> ..... | IV-10 |
| 4.4.     | Fase Konstruksi .....   | IV-10 |
| 4.4.1.   | Diagram Kelas Keseluruhan.....  | IV-11 |
| 4.4.2.   | Implementasi Kelas .....  | IV-12 |
| 4.4.3.   | Implementasi Antarmuka .....  | IV-13 |
| 4.5.     | Fase Transisi.....  | IV-13 |
| 4.5.1.   | Pengujian Perangkat Lunak.....  | IV-13 |
| 4.5.1.1. | Rencana Pengujian.....  | IV-13 |
| 4.5.1.2. | Kasus Uji.....  | IV-15 |
| 4.5.1.3. | Hasil Pengujian <i>Use Case</i> .....                                       | IV-17 |
| 4.6.     | Kesimpulan.....   | IV-19 |

## **BAB V PENGEMBANGAN PERANGKAT LUNAK**

|        |   |     |
|--------|---|-----|
| 5.1.   | Pendahuluan .....                             | V-1 |
| 5.2.   | Data Hasil Percobaan .....                    | V-1 |
| 5.2.1. | Konfigurasi Percobaan .....                   | V-1 |
| 5.2.2. | Data Hasil Percobaan <i>Naïve Bayes</i> ..... | V-1 |
| 5.3.   | Analisis Hasil Penelitian .....               | V-4 |
| 5.4.   | Kesimpulan.....                               | V-5 |

**BAB VI KESIMPULAN DAN SARAN**

|                        |      |
|------------------------|------|
| 6.1. Pendahuluan ..... | VI-1 |
| 6.2. Kesimpulan.....   | VI-1 |
| 6.3. Saran.....        | VI-2 |

|                             |            |
|-----------------------------|------------|
| <b>DAFTAR PUSTAKA .....</b> | <b>xix</b> |
|-----------------------------|------------|

|                       |            |
|-----------------------|------------|
| <b>LAMPIRAN .....</b> | <b>L-1</b> |
|-----------------------|------------|

## DAFTAR GAMBAR

|              |  | Halaman |
|--------------|--|---------|
| Gambar II-1  | Struktur pesan dari sudut pandang ekstraksi fitur.....   | II-2    |
| Gambar II-2  | Contoh visual dari struktur pesan.....   | II-2    |
| Gambar II-3  | Contoh visual pesan sampah (spam).....   | II-3    |
| Gambar II-4  | Contoh visual pesan <i>non-spam</i> .....  | II-4    |
| Gambar II-5  | Ilustrasi <i>iterative</i> RUP.....  | II-8    |
| Gambar III-1 | Proses <i>Pre-processing</i> .....   | III-3   |
| Gambar III-2 | Proses Klasifikasi Menggunakan Naive Bayes.....  | III-6   |
| Gambar III-3 | Penjadwalan untuk Tahap Menentukan Ruang Lingkup dan Unit Penelitian.....  | III-26  |
| Gambar III-4 | Penjadwalan untuk Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian dan Menentukan Kriteria Pengujian..... | III-27  |
| Gambar III-5 | Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Insepsi.....                  | III-27  |
| Gambar III-6 | Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Elaborasi...                  | III-28  |
| Gambar III-7 | Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Konstruksi..                  | III-28  |
| Gambar III-8 | Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Transisi...                   | III-29  |
| Gambar III-9 | Penjadwalan untuk Tahap Melakukan Pengujian Penelitian, Analisa Hasil Pengujian Penelitian dan Membuat Kesimpulan..... | III-29  |
| Gambar IV-1  | Diagram <i>Use Case</i> Pendeteksian Spam.....   | IV-3    |
| Gambar IV-2  | Kelas Analisis Mendeteksi Email dengan Naive Bayes.....  | IV-6    |
| Gambar IV-3  | Perancangan Antarmuka <i>MainFrame</i> .....   | IV-8    |
| Gambar IV-4  | Diagram <i>Sequence</i> Mendeteksi Email dengan <i>Naive Bayes</i>   | IV-9    |

|             |   |       |
|-------------|---|-------|
| Gambar IV-5 | Kelas Diagram Keseluruhan Pendeteksian Spam.....                                  | IV-11 |
| Gambar IV-6 | Interface MainFrame.....  | IV-13 |
| Gambar IV-7 | Hasil Pengujian <i>Use Case</i> Mendeteksi Email Dengan Naive Bayes (UC-101)..... | IV-17 |
| Gambar IV-8 | Hasil Pengujian <i>Use Case</i> Mendeteksi Email Dengan Naive Bayes (UC-102)..... | IV-18 |
| Gambar IV-9 | Hasil Pengujian <i>Use Case</i> Mendeteksi Email Dengan Naive Bayes (UC-103)..... | IV-19 |
| Gambar V-1  | Grafik Akurasi metode <i>Naive Bayes</i> .....                                    | V-4   |

## DAFTAR TABEL

|             |  | Halaman |
|-------------|--|---------|
| Table III-1 | Contoh Proses <i>Tokenizing</i> .....  | III-4   |
| Tabel III-2 | Contoh Proses <i>Case Folding</i> .....  | III-4   |
| Tabel III-3 | Contoh Frekuensi <i>Token</i> dalam <i>Dataset</i> .....   | III-7   |
| Tabel III-4 | Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure (WBS)</i> .....                        | III-22  |
| Tabel IV-1  | Definisi Aktor.....  | IV-4    |
| Tabel IV-2  | Definisi <i>Use Case</i> .....   | IV-4    |
| Tabel IV-3  | Skenario <i>Use Case</i> Mendeteksi <i>Email</i> Dengan <i>Naive Bayes</i> .....                       | IV-5    |
| Tabel IV-4  | Daftar Implementasi Kelas.....   | IV-12   |
| Tabel IV-5  | <i>Interface MainFrame</i> .....   | IV-14   |
| Tabel IV-6  | Rencana Pengujian <i>Use Case</i> Mendeteksi Email Dengan <i>Naive Bayes</i> .....                     | IV-15   |
| Tabel V-1   | Hasil pengujian pendeteksian <i>spam</i> dengan <i>Naive Bayes</i> (Klasifikasi <i>non-spam</i> )..... | V-1     |



|           |  |     |
|-----------|--|-----|
| Tabel V-2 | Hasil pengujian pendeteksian <i>spam</i> dengan <i>Naive Bayes</i><br>( <i>spam</i> ).....       | V-3 |
| Tabel V-3 | Hasil pengujian pendeteksian <i>spam</i> dengan <i>Naive Bayes</i><br>(campur).....              | V-3 |
| Tabel V-4 | Hasil pengujian pendeteksian <i>spam</i> dengan <i>Naive Bayes</i><br>( <i>Real email</i> )..... | V-3 |

## DAFTAR LAMPIRAN

|                             | Halaman |
|-----------------------------|---------|
| LAMPIRAN KELAS DIAGRAM..... | L-1     |
| LAMPIRAN SAMPEL DATA.....   | L-3     |
| LAMPIRAN KODE PROGRAM.....  | L-6     |

# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab I akan menjelaskan gambaran secara umum penelitian. Penjelasan dimulai dari latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian hingga sistematika penulisan.

### **1.2 Latar Belakang Masalah**

Perkembangan internet secara bertahap menjadi bagian yang penting dalam kehidupan sehari-hari setiap orang. Penggunaan internet diperkirakan akan terus berkembang dan *email* telah menjadi alat yang kuat ditujukan untuk pertukaran ide dan informasi. Banyak hal yang dapat dirasakan dalam berkomunikasi, terutama *email* yang merupakan bentuk komunikasi termurah dan mampu menjangkau jutaan pengguna dengan waktu yang singkat. Menurut Idris *et.al* (2013), yang menjadikan *email* penting hingga saat ini adalah efektif, sederhana dan tersedia untuk semua pengguna komputer.

Seiring dengan pertumbuhan Internet dan *email*, telah terjadi pertumbuhan *spam* yang dramatis dalam beberapa tahun terakhir. Oxford Online Dictionary mendeskripsikan *spam* sebagai “pesan tidak relevan atau tidak diinginkan yang dikirim melalui Internet, biasanya untuk tujuan iklan, *phishing*, menyebarkan *malware*, dll.” *Spam* juga dikenal sebagai *email* komersial.

Masalah yang ditimbulkan oleh *spam* telah meningkat setiap tahunnya. Sebuah statistik tahun 2011 memaparkan bahwa 40% dari semua *email spam* atau tepatnya 15.4 triliun spam email tersebar setiap harinya dan berdampak terhadap pengguna internet sehingga menyebabkan kerugian hingga 355 juta dolar per tahun (Awad dan Elsoufi, 2011). *Email* menyediakan cara yang terbaik dalam mengirimkan jutaan iklan bebas biaya oleh pengirim. Namun, fakta ini secara luas dimanfaatkan oleh beberapa individu dan organisasi. Akibatnya, kotak surat elektronik jutaan orang dipenuhi dengan semua email yang tidak diinginkan yang juga diketahui sebagai “*spam*” atau “*junk mail*”.

Namun, pengguna tidak dapat menghindari masalah serius dalam upaya menangani sejumlah besar spam. Jika tidak ada aktivitas *anti-spam*, dampak besar dari volume *spam* akan membanjiri sistem jaringan, membunuh produktivitas karyawan, mencuri *bandwidth*, dan mengurangi kontrol keamanan yang dapat merusak sistem komputer atau perangkat seluler.

Cara yang paling sederhana untuk mengatasi *spam* adalah dengan memblokir *email* yang datang dari alamat IP tertentu atau dengan memfilter judul subjek tertentu. Metode yang paling umum yang disebut sebagai pendeteksian *anti-spam* bekerja dengan memisahkan atau mengelompokkan isi *email* kemudian mempelajari sejumlah *email spam* yang telah terklasifikasi. Proses ini dapat mencegah *spam* masuk kedalam kotak pesan dan memperingatkan penerima tentang *spam* tersebut dikemudian harinya. Aktor yang menyebarkan *email* yang mengandung *spam* disebut sebagai *spammer*. Saat ini *spammer* menggunakan metode yang unik untuk dapat melewati *filter* dengan mengirimkan *spam* dengan

alamat acak atau memasukkan karakter acak pada awal dan akhir subjek *email* untuk melewati mesin pendeteksi *anti-spam*. Perilaku ini menjadi alasan mengapa masalah *spam* tampaknya masih bertahan dan dengan teknik terancang masa kini pun tidak memberikan cukup bukti dalam mengatasi masalah ini.

Saat ini, banyak penelitian yang telah dilakukan untuk mendeteksi *spam* dengan menggunakan *Bayesian classifier*. Marsono et al., (2009) telah mendemonstrasikan bahwa konten klasifikasi *Naive Bayes* dapat diadaptasi untuk *prosesing layer-3*, tanpa memerlukan perakitan ulang.

Metode *Naive Bayes* dijadikan subjek karena metode tersebut menggunakan probabilitas dan statistik dalam pendeteksian. Menurut Awad dan Elsoufi (2011) *machine learning* telah dipelajari secara luas dan terdapat banyak metode yang dapat digunakan dalam penyaringan *email*. Salah-satu metode yang disebut adalah *Naive Bayes*.

### 1.3 Rumusan Masalah

Penelitian bidang pendeteksi *spam* telah banyak dilakukan dan terus mengalami perkembangan. Metode-metode baru terus dicoba untuk mendapatkan hasil akurasi yang maksimal. Dalam penelitian Awad dan Elsoufi (2011) *Naive Bayes* mendapatkan hasil akurasi yang tinggi sejumlah 99.46%. Pembuktian akurasi metode *Naive Bayes* terhadap sistem pendeteksi *spam* menjadi alasan penelitian ini. Setiap email memiliki isi pesan yang beragam sehingga kemungkinan kesalahan pendeteksi dapat terjadi. Adapun pertanyaan utama pada penelitian adalah “Bagaimana tingkat akurasi dari metode *Naive Bayes* dalam menyaring *spam*?”

Pertanyaan pendukung penelitian ini adalah :

- 1) Bagaimana implementasi metode pendeteksian *email spam* dan *email non-spam* dengan metode pengklasifikasi *Bayesian*.

#### **1.4 Tujuan Penelitian**

Tujuan penelitian ini adalah sebagai berikut:

1. Mengembangkan perangkat lunak komputer untuk pendeteksian menggunakan pengklasifikasi *Bayesian* untuk membaca fitur yang telah disediakan.
2. Memperoleh tingkat akurasi dalam mendeteksi *spam email* menggunakan metode pengklasifikasi *Bayesian*.

#### **1.5 Manfaat Penelitian**

Manfaat penelitian ini adalah:

1. Mendapatkan hasil dari klasifikasi *email spam* dan *email non-spam*.
2. Membuktikan akurasi dari algoritma *Bayesian* yang diterapkan ke dalam perangkat lunak yang akan dibangun.
3. Hasil penelitian dapat digunakan sebagai acuan untuk penelitian-penelitian pendeteksi spam email selanjutnya.

## 1.6 Batasan Masalah

Batasan masalah dalam penelitian tugas akhir ini adalah:

1. Inputan *email* di dalam aplikasi berupa teks. Tidak meliputi gambar dan *attachment* multimedia di dalamnya.
2. Pesan *email* diklasifikasikan kedalam *spam* dan *non-spam*.
3. *Email* yang di-*training* berbahasa Inggris.
4. Urutan pembelajarannya akan menggunakan sekumpulan *dataset offline* tangan kedua (sekunder) dari *website* pada alamat berikut <http://www2.aueb.gr/users/ion/data/enron-spam/>
5. *Email* yang di *training* berjumlah sebanyak 200 *email* yang terdiri dari 100 *email* untuk *training* data dan 100 *email* untuk pengujian.
6. Penelitian menitikberatkan terhadap akurasi dari metode yang diuji.
7. Penelitian yang dimaksud akan dikembangkan menggunakan bahasa pemrograman Java.

## 1.7 Sistematika Penulisan

Sistematika penulisan laporan tugas akhir ini adalah sebagai berikut :

### **BAB I. PENDAHULUAN**

Bab ke 1 berisi penjelasan mengenai latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, metodologi penelitian, metode pengembangan perangkat lunak dan sistematika penulisan.

## **BAB II. KAJIAN LITERATUR**

Bab ke 2 menguraikan dasar-dasar teori yang digunakan pada penelitian, seperti *Electronic Mail, Pre-Processing, Bayesian Classifier, Jaringan Syaraf Tiruan*, pengembangan perangkat lunak, desain model, bahasa pemrograman Java.

## **BAB III. METODOLOGI PENELITIAN**

Bab ke 3 berisi analisis serta perancangan terhadap penggunaan algoritma *Bayesian* dalam melakukan pelatihan dan pengenalan pada sampel *email* yang dikembangkan untuk pembandingan dalam perangkat lunak pendeteksi *spam*.

## **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Bab ke 4 membahas mengenai lingkungan implementasi algoritma *Bayesian* dalam melakukan pelatihan dan pengenalan pada sampel *email* serta implementasi program dan pengujian.

## **BAB V. HASIL DAN ANALISIS PENELITIAN**

Bab ke 5 berisi kesimpulan dari semua uraian-uraian pada bab - bab sebelumnya dan saran untuk penelitian selanjutnya.

## **BAB VI. KESIMPULAN DAN SARAN**

Bab ke 6 berisi kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam pembandingan pendeteksi *spam* ini.



## 1.8 Kesimpulan

Penelitian yang bertajuk Implementasi *Naive Bayes* untuk pendeteksian *spam email* berlatar belakang atas rasa ingin tahu peneliti terhadap tingkat kualitas metode *Naive Bayes* dalam mendeteksi *spam*. Dibatasi *dataset offline*, penelitian diharapkan menunjukkan akurasi dari metode yang digunakan sehingga menghasilkan bahan kajian untuk metode terbaik.

## DAFTAR PUSTAKA

- Ambler, S.W. (2005). A Manager's Introduction to the Rational Unified Process (RUP).
- Bonczek, R. H., Holsapple, C. W., & Whinston, A. B. (1980). FUTURE DIRECTIONS FOR DEVELOPING DECISION SUPPORT SYSTEMS\*. *Decision Sciences*, 11(4), 616-631.
- Carvalho, J. A. B. (2006). Rational and Efficient use of Energy in Industrial Electrical Installations. In *Universities Power Engineering Conference, 2006. UPEC'06. Proceedings of the 41st International* (Vol. 3, pp. 951-955). IEEE.
- Davis, J., Srivina, A., & Sundaram, D. (2000). Implementing Decision Support Systems : Methods, Techniques, and Tools. MC Graw Hill.
- Dicks, L. V., Walsh, J. C., & Sutherland, W. J. (2014). Organising evidence for environmental management decisions: a '4S'hierarchy. *Trends in ecology & evolution*, 29(11), 607-613.
- Hwang, C.L.; Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications*. New York: Springer-Verlag.
- Kosasi, S. (2002). Sistem Penunjang Keputusan (Decision Support System). *Pontianak: Departemen Pendidikan Nasional*.
- Knoernschild, K. (2002). *Java design: objects, UML, and process*. Addison-Wesley Professional.
- Kusumadewi, S., Hartati, S., Harjoko, A., Wardoyo, R. (2006). *Fuzzy Multi-Attribute Decision Making (Fuzzy MADM)*. Graha Ilmu, Yogyakarta.
- Kruchten, P. (2001). Agility with the RUP. *Cutter IT Journal*, 14(12), 27-33.

- Pangeran, M. (2010). Sistem Pendukung Keputusan Selesi Penerima Beasiswa Dengan Metode AHP dan TOPSIS. Universitas Sumatera Utara
- Perdana, Nuri Guntur and Widodo, Tri (2013) *Sistem Pendukung Keputusan Pemberian Beasiswa Kepada Peserta Didik Baru Menggunakan Metode TOPSIS*. Semantik 2013, 3 (1). pp. 265-272. ISSN 979-26-0266-6
- Pramuditha, A. N., Suyono, H., Yudaningtyas, E. (2015). Penggunaan Algoritma *Multi Criteria Decision Making* dengan Metode TOPSIS dalam Penempatan Karyawan. *Jurnal EECCIS* (Vol. 9, No. 1)
- Sprague Jr, R. H. (1980). A framework for the development of decision support systems. *MIS quarterly*, 1-26.
- Sudrajat, F.A. & Nuryana, I.K.D. (2014). Sistem Pendukung Keputusan Pemilihan Otobus Menggunakan Metode TOPSIS Berbasis Web Dengan YII Framework. *Jurnal Managemen Informatika*. (Vol. 01 No. 01. 9-15).
- Turban, E., Aronson, J., & Liang, T. P. (2005). *Decision Support Systems and Intelligent Systems 7* "" Edition. Pearson Prentice Hall.
- Wang, W., Fan, K., Su, Y., & Liang, S. (2008, October). Air combat effectiveness assessment of military aircraft using a fuzzy AHP and TOPSIS methodology. In *System Simulation and Scientific Computing, 2008. ICSC 2008. Asia Simulation Conference-7th International Conference on* (pp. 655-662). IEEE.