

**PENDETEKSIAN EMAIL *SPAM* DAN *NON-SPAM* DENGAN METODE
KLASIFIKASI NAIVE BAYES**

Diajukan Sebagai Syarat untuk Menyelesaikan
Pendidikan Program Strata-1 pada
Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI



Oleh:

Abu Hanifah Warga Dalam

09021381520049

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMUKOMPUTER UNIVERSITAS SRIWIJAYA**

2020

**PENDETEKSIAN EMAIL *SPAM* DAN *NON-SPAM* DENGAN METODE
KLASIFIKASI NAIVE BAYES**

Diajukan Sebagai Syarat untuk Menyelesaikan
Pendidikan Program Strata-1 pada
Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI



Oleh:

Abu Hanifah Warga Dalam

09021381520049

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMUKOMPUTER UNIVERSITAS SRIWIJAYA**

2020

LEMBAR PENGESAHAN TUGAS AKHIR

***PENDETEKSIAN EMAIL SPAM DAN NON-SPAM DENGAN METODE
KLASIFIKASI NAIVE BAYES***

Oleh:

Abu Hanifah Warga Dalam
NIM : 09021381520049

Palembang, Juli 2020

Pembimbing I,



Drs. Megah Mulya, M.T
NIP. 196602202006041001

Pembimbing II,



Kanda Januar Miraswan, M.T
NIP. 199001092019031012

Mengetahui,
Ketua Jurusan Teknik Informatika,


Rifkie Primartha, MT
NIP. 197706012009121004

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Jum'at tanggal 13 Maret 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Abu Hanifah Warga Dalam
NIM : 09021381520049
Judul : Pendeteksian Email *Spam* dan *Non-spam* dengan Metode Klasifikasi Naive Bayes

1. Pembimbing I

Drs. Megah Mulya, M.T
NIP. 196602202006041001

2. Pembimbing II

Kanda Januar Miraswan, M.T
NIP. 199001092019031012

3. Penguji I

Yunita, M.CS
NIP. 198306062015042000

4. Penguji II

Mastura Diana Marieska, M.T
NIP. 198503212018032001

Mengetahui,
Ketua Jurusan Teknik Informatika

Rifkie Primartha, MT
NIP. 197706012009121004



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Abu Hanifah Warga Dalam
NIM : 09021381520049
Program Studi : Teknik Informatika
Judul Skripsi : Pendeteksian Email *Spam* dan *Non-spam* dengan Metode Klasifikasi Naive Bayes
Hasil Pengecekan Software *iThenticate/Turnitin* : 17 %

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek 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, Juli 2020



Abu Hanifah Warga Dalam
NIM. 09021381520049

MOTTO DAN PERSEMBAHAN

MOTTO:

- If you regret your past, then make your future better.
- Life is not a video game, we only have one life left. So do the best for your life.

Kupersembahkan karya tulis ini kepada:

- *Kedua orang tuaku tercinta dan adik-adikku tersayang*
- *Keluarga besarku*
- *Teman-teman Informatika 2015*
- *Fakultas Ilmu Komputer*
- *Universitas Sriwijaya*

DETECTION OF SPAM AND NON-SPAM E-MAIL USING THE NAIVE BAYES
CLASSIFICATION METHOD

By:

Abu Hanifah Warga Dalam

09021381520049

ABSTRACT

Message exchanging in the form of e-mail or electronic mail has become one of the most important things in this information age. Every e-mail has different message content from personal messages to commercial messages. However, as the development of services and technology keep growing, the number of spam messages continues to grow rapidly. This problem can be controlled with a spam detection system. Spam detection system in e-mail is a system that groups or separates messages in e-mail. In this research, there are several processes carried out to detect spam, starting from pre-processing with tokenizing and case folding. After going through the preprocessing stages, training and testing will be conducted using the Naive Bayes method. This study aims to measure and prove the level of accuracy in detecting spam and non-spam emails by the Naive Bayes method.

Keywords: Spam Detection, E-mail, Tokenizing, Case Folding, Naive Bayes.

Palembang, July 2020

Supervisor II,

Supervisor I,



Drs. Megah Mulya, M.T

NIP. 196602202006041001



Kanda Januar Miraswan, M.T

NIP. 199001092019031012

Approve,

Head of the Informatics Engineering Department,



Rifkie Primartha, MT

NIP. 197706012009121004

PENDETEKSIAN EMAIL SPAM DAN NON-SPAM DENGAN METODE
KLASIFIKASI NAIVE BAYES

Oleh:

Abu Hanifah Warga Dalam

09021381520049

ABSTRAK

Media pertukaran pesan dalam bentuk email atau *electronic mail* telah menjadi salah satu hal terpenting pada era informasi ini. Setiap email memiliki isi pesan yang berbeda-beda dari pesan yang bersifat pribadi hingga pesan yang bersifat komersil. Akan tetapi, seakan berkembangnya layanan dan teknologi, jumlah pesan spam tetap terus bertambah pesat. Masalah ini dapat dikontrol dengan sebuah sistem pendeteksian spam. Sistem pendeteksi spam pada email adalah sistem yang mengelompokkan atau memisahkan pesan dalam email. Dalam penelitian ini, ada beberapa proses yang dilakukan untuk mendeteksi spam yaitu dimulai dari praproses dengan *tokenizing* dan *case folding*. Setelah melewati tahapan praproses, kemudian akan dilakukan pelatihan dan pengujian dengan metode Naive Bayes. Penelitian ini bertujuan untuk mengukur serta membuktikan tingkat akurasi dalam mendeteksi email *spam* dan *non-spam* dengan metode Naive Bayes.

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

Palembang, Juli 2020

Pembimbing I,



Drs. Megah Mulya, M.T

NIP. 196602202006041001

Pembimbing II,



Kanda Januar Miraswan, M.T

NIP. 199001092019031012

Mengetahui,

Ketua Jurusan Teknik Informatika,



Rifkie Primartha, MT

NIP. 197706012009121004

KATA PENGANTAR

Puji dan syukur kepada Allah atas berkat serta rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik serta memuaskan. Tugas akhir ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Orang tuaku, M. Andra Adityawarman dan Eka Desianti, adik-adikku, Fatimah Azzahra, Dzakiyah Al-Munawaroh, dan Sabral Jamilah, serta seluruh keluarga besarku yang selalu mendoakan serta memberikan dukungan baik moral maupun materil. Terkhususnya untuk Almarhumah nenekku, Yuhana yang terus memotivasi dan mendoakan penulis hingga akhir hayat beliau.
2. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Bapak Rifkie Primartha, M.T selaku Ketua Jurusan Teknik Informatika, dan Ibu Alvi Syahrini Utami, M.Kom, selaku Sekretaris Jurusan Teknik Informatika.
3. Bapak Megah Mulya, M.T, selaku dosen pembimbing I serta Bapak Kanda Januar M, M.T selaku pembimbing II yang telah membimbing dan memberi arahan kepada penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir ini.
4. Ibu Novi Yusliani, M.T selaku dosen pembimbing akademik, yang telah membimbing memberi arahan serta memberi motivasi kepada penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir ini.

5. Ibu Yunita, M.Sc selaku dosen penguji I dan Ibu Mastura Diana Marieska, S.T, M.T selaku dosen penguji II yang telah memberikan masukan, kritik dan saran dalam proses pengerjaan Tugas Akhir.
6. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya
7. Mbak Wiwin, dan seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Teman-teman SMA ku (Fariz, Gad, Harvi, Erlangga, Reza) yang selalu menyemangati untuk menyelesaikan skripsi serta selalu mendengarkan keluhan penulis dan memberi saran dalam menyelesaikan skripsi ini.
9. Teman-teman seperjuangan dalam mengerjakan Tugas Akhir ini (Agus, Adi, Ajrul, Gheddi, Ilham, Winto) yang telah mendampingi penulis dalam proses penyelesaian tugas akhir ini.
10. Teman-teman IFBIL A 2015 dan seluruh teman-teman Teknik Informatika Universitas Sriwijaya.

Penulis menyadari dalam penyusunan Tugas Akhir ini jauh dari kata sempurna dan masih banyak kekurangan disebabkan oleh keterbatasan pengetahuan dan pengalaman, oleh karena itu penulis mengharapkan kritik dan saran yang membangun dari semua pihak untuk penyempurnaan tugas akhir ini untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, Februari 2020

Abu Hanifah Warga Dalam

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvii

BAB I PENDAHULUAN

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

BAB II KAJIAN LITERATUR

2.1 Pendahuluan.....	II-1
----------------------	------

2.2 Landasan Teori.....	II-1
2.2.1 <i>Email</i>	II-1
2.2.2 <i>Data Mining</i>	II-5
2.2.3 <i>Pre-processing</i>	II-6
2.2.4 Klasifikasi Naive Bayes	II-8
2.2.5 Rational Unified Process (RUP).....	II-10
2.3 Penelitian Lain yang Relevan.....	II-14
2.3.1 Analisis Spam dengan Menggunakan Naive Bayes	II-14
2.3.2 <i>An Implementation of Naive Bayes Method on Email Spam And Non-spam Detection</i>	II-15
2.4 Kesimpulan	II-15

BAB III METODOLOGI PENELITIAN

3.1 Pendahuluan	III-1
3.2 Metode Pengumpulan Data.....	III-1
3.3 Tahapan Penelitian	III-2
3.3.1 Menetapkan Kerangka Kerja / Framework	III-2
3.3.2 Menetapkan Kriteria Pengujian	III-4
3.3.3 Menetapkan Format Data Pengujian.....	III-5
3.3.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian.....	III-6
3.3.5 Melakukan Pengujian Penelitian	III-6
3.3.6 Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan Penelitian	III-7
3.4 Metode Pengembangan Perangkat Lunak.....	III-7
3.4.1 <i>Rational Unified Process (RUP)</i>	III-8
3.5 Manajemen Proyek Penelitian.....	III-11
3.6 Kesimpulan	III-27

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1 Pendahuluan	IV-1
4.2 <i>Rational Unified Process</i> (RUP)	IV-1
4.2.1 Fase Insepsi	IV-1
4.2.2 Fase Elaborasi	IV-27
4.2.3 Fase Konstruksi	IV-39
4.2.4 Fase Transisi	IV-48
4.3 Kesimpulan	IV-55

BAB V HASIL DAN ANALISIS PENELITIAN

5.1 Pendahuluan	V-1
5.2 Data Hasil Percobaan/Penelitian	V-1
5.2.1 Konfigurasi Percobaan	V-1
5.2.2 Data Hasil Konfigurasi Percobaan Naive Bayes	V-2
5.3 Analisis Hasil Penelitian	V-5
5.4 Kesimpulan	V-7

BAB VI KESIMPULAN DAN SARAN

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

DAFTAR PUSTAKA	xviii
-----------------------------	-------

LAMPIRAN	xx
-----------------------	----

DAFTAR TABEL

	Halaman
Tabel II-1 Case Folding	II-6
Tabel II-2 Tokenizing	II-7
Tabel III-1 Rancangan tabel jenis email pada masing-masing email	III-5
Tabel III-2 Rancangan Tabel Hasil Klasifikasi <i>Naive Bayes</i>	III-7
Tabel III-3 Tabel Penjadwalan Penelitian dalam Bentuk WBS	III-12
Tabel IV-1 Kebutuhan Fungsional	IV-3
Tabel IV-2 Kebutuhan Non-Fungsional	IV-4
Tabel IV-3 Teks <i>Email</i>	IV-6
Tabel IV-4 Proses <i>Tokenizing</i>	IV-6
Tabel IV-5 Proses <i>Case Folding</i>	IV-6
Tabel IV-6 Contoh Frekuensi Token dalam Dataset	IV-7
Tabel IV-7 Perhitungan Probabilitas masing-masing Token	IV-8
Tabel IV-8 Perhitungan Probabilitas Email Spam.....	IV-20
Tabel IV-9 Definisi Aktor	IV-21
Tabel IV-10 Definisi Use Case.....	IV-22
Tabel IV-11 <i>Use Case</i> Memilih <i>File</i> Data Latih	IV-23
Tabel IV-12 <i>Use Case</i> Memilih <i>File</i> Data Uji	IV-24
Tabel IV-13 <i>Use Case</i> Melakukan klasifikasi <i>Naive Bayes</i>	IV-25
Tabel IV-14 <i>Use Case</i> Melihat <i>Token</i>	IV-26
Tabel IV-15 Kebutuhan Sistem	IV-31
Tabel IV-16 Implementasi Kelas.....	IV-43
Tabel IV-17 Rencana Pengujian <i>Use Case</i> Memilih File Data Latih.....	IV-49
Tabel IV-18 Rencana Pengujian <i>Use Case</i> Memilih File Data Uji.....	IV-49
Tabel IV-19 Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi	

Naive Bayes.....	IV-50
Tabel IV-20 Rencana Pengujian <i>Use Case</i> Melihat <i>Token</i>	IV-50
Tabel IV-21 Pengujian <i>Use Case</i> Memilih File Data Latih	IV-51
Tabel IV-22 Pengujian <i>Use Case</i> Memilih File Data Uji	IV-52
Tabel IV-23 Pengujian <i>Use Case</i> Melakukan Klasifikasi Naive Bayes	IV-53
Tabel IV-24 Pengujian <i>Use Case</i> Melihat <i>Token</i>	IV-54
Tabel V-1 Hasil Pengujian Pendeteksian Email dengan Naive Bayes	V-2
Tabel V-2 <i>Confusion Matrix</i> Pengujian Pendeteksian Email dengan Naive Bayes	V-4
Tabel V-3 Hasil Pengujian Pendeteksian Email <i>Real</i> dengan Naive Bayes ..	V-4
Tabel V-2 <i>Confusion Matrix</i> Pengujian Pendeteksian Email <i>Real</i> dengan Naive Bayes	V-5

DAFTAR GAMBAR

	Halaman
Gambar II-1 Arsitektur Email	II-2
Gambar II-2 Contoh dari Pesan <i>Spam</i>	II-4
Gambar II-3 Contoh dari Pesan <i>Non-spam</i>	II-5
Gambar II-4 Arsitektur <i>Rational Unified Process</i>	II-11
Gambar III-1 Diagram Tahapan Proses Perangkat Lunak.....	III-3
Gambar III-2 Diagram Tahapan Penelitian	III-6
Gambar III-3 Tahap Menentukan Ruang Lingkup dan Unit Penelitian.....	III-22
Gambar III-4 Tahap Menentukan Dasar Teori Berkaitan dengan Penelitian .	III-22
Gambar III-5 Tahap Menentukan Kriteria Pengujian	III-23
Gambar III-6 Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Inception.....	III-23
Gambar III-7 Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Elaboration	III-24
Gambar III-8 Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Construction	III-24
Gambar III-9 Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Transition	III-25
Gambar III-10 Tahap Melakukan Pengujian Penelitian	III-25
Gambar III-11 Tahap Membuat Analisa Hasil Pengujian dan Kesimpulan	III-26
Gambar IV-1 Diagram Use Case	IV-21
Gambar IV-2 Perancangan Antarmuka MainFrame	IV-29
Gambar IV-3 Perancangan Antarmuka ClassificationFrame Naive Bayes...	IV-30
Gambar IV-4 Perancangan Antarmuka TokenFrame.....	IV-30
Gambar IV-5 Diagram Aktivitas Memilih Data Latih	IV-33
Gambar IV-6 Diagram Aktivitas Memilih Data Uji	IV-34

Gambar IV-7	Diagram Aktivitas Melakukan Klasifikasi Naive Bayes	IV-35
Gambar IV-8	Diagram Aktivitas Melihat <i>Token</i>	IV-36
Gambar IV-9	Diagram <i>Sequence</i> Memilih <i>File</i> Data Latih.....	IV-37
Gambar IV-10	Diagram <i>Sequence</i> Memilih <i>File</i> Data Uji.....	IV-37
Gambar IV-11	Diagram <i>Sequence</i> Melakukan Klasifikasi <i>Naive Bayes</i>	IV-38
Gambar IV-12	Diagram <i>Sequence</i> Melihat <i>Token</i>	IV-39
Gambar IV-13	Diagram Kelas	IV-40
Gambar IV-14	Kelas Analisis Memilih Data Latih	IV-41
Gambar IV-15	Kelas Analisis memilih Data Uji	IV-41
Gambar IV-16	Kelas Analisis Melakukan Klasifikasi <i>Naive Bayes</i>	IV-42
Gambar IV-17	Kelas Analisis Melihat <i>Token</i>	IV-42
Gambar IV-18	Antarmuka <i>MainFrame</i>	IV-46
Gambar IV-19	Antarmuka <i>ClassificationFrame</i>	IV-47
Gambar IV-20	Antarmuka <i>TokenFrame</i>	IV-48
Gambar V-1	Grafik Akurasi metode Naive Bayes	V-6

DAFTAR LAMPIRAN

Lampiran 1. Sampel Data.....	L-1
Lampiran 2. <i>Source Code</i> Program.....	L-9

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah/ruang lingkup, dan sistematika penulisan pada penelitian.

1.2 Latar Belakang

Penggunaan internet semakin menjadi sebuah bagian dari kehidupan sehari-hari kita. Penggunaan internet yang terus berkembang dan dengan adanya *email* yang menjadi sebuah alat yang mampu untuk media bertukar pesan, baik untuk keperluan komersil ataupun keperluan pribadi. Seiring dengan berkembang pesatnya internet dan *email*, perkembangan *email spam* juga berkembang pesat.

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 (Ghani & Subekti, 2018).

Akan tetapi, dengan berkembangnya serta banyaknya layanan dan teknologi *anti-spam*, jumlah pesan *spam* tetap bertambah secara pesat (Mishra & S. Thakur, 2013). 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* (Rusland, Wahid, Kasim, & Hafit, 2017). 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 kumpulan *set* pesan *spam* dan *non-spam* 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.

Penelitian ini akan membandingkan algoritma Naive Bayes untuk mendeteksi *spam email* yang tujuannya untuk melihat sebuah akurasi dan efisiensi dalam sistem pendeteksian *email spam*. Penulis memilih metode tersebut untuk menelusuri konsep dibalik pendeteksian email spam menggunakan Naive Bayes.

1.3 Rumusan Masalah

Rumusan masalah yang mendasari penelitian ini adalah:

Bagaimana cara mengimplementasikan algoritma Naive Bayes pada pendeteksian email spam?

1.4 Tujuan Penelitian

Tujuan Penelitian ini adalah:

1. Mengimplementasikan algoritma Naive Bayes pada pendeteksian *email spam*.
2. Mengetahui tingkat akurasi dari algoritma Naive Bayes pada pendeteksian *email spam*.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah membantu mengenali perbedaan email *spam* dan email *non-spam*. Serta, hasil dari penelitian dapat digunakan sebagai referensi untuk penelitian metode pendeteksian *spam* lebih lanjut.

1.6 Batasan Masalah

Batasan masalah yang ditetapkan dalam penelitian ini adalah:

1. Masukan email yang ada hanya berupa teks. Tidak termasuk gambar dan berbagai lampiran lainnya.
2. Email dikategorikan sebagai *spam* dan *non-spam*.
3. Email yang digunakan dalam Bahasa Inggris.
4. Bagian email yang akan digunakan adalah Subjek dan Isi.
5. Email yang akan digunakan sebagai data latih adalah 200 email yang diantaranya 70% email sebagai data latih dan 30% email sebagai data uji.
6. Fokus dari penelitian ini adalah perhitungan tingkat akurasi dari metode yang akan diterapkan.

1.7 Sistematika Penulisan

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah/ruang lingkup, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian, seperti definisi Electronic Mail, Spam, Non-Spam, Machine Learning, Pre-processing, Naive Bayes Classifier, Algoritma Random Forest.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan dibahas mengenai tahapan dan komponen pendukung yang terdapat pada penelitian ini. Mulai dari unit penelitian, sumber data, jenis data, pengumpulan data, metode pengembangan perangkat lunak *Rational Unified Process* (RUP), dan masing-masing rencana tahapan penelitian dengan rinci.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini akan membahas perancangan perangkat lunak yang akan dibangun pada penelitian ini.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini akan menampilkan hasil pengujian/penelitian berdasarkan langkah langkah yang telah direncanakan. Analisis diberikan sebagai basis dari kesimpulan yang diambil dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam pengembangan penelitian ini selanjutnya.

1.8 Kesimpulan

Beberapa email spam tidak ditandai sebagai spam karena email *filtering* tidak mendeteksi bahwa email itu adalah email spam. Hal ini menjadi sebuah ancaman bukan hanya di dunia internet saja, tetapi untuk masyarakat juga akan berpengaruh. Oleh karena itu, penelitian ini akan menggunakan algoritma Naive

Bayes untuk mendeteksi spam email yang tujuannya untuk melihat sebuah akurasi dalam sistem pendeteksian email spam.

DAFTAR PUSTAKA

- Bayati, M. A., & Jabbar, S. F. (2015). Developing a Spam Email Detector. *International Journal of Engineering and Innovative Technology (IJEIT)*, 5(2), 16–21.
- Bustami. (2014). Penerapan Algoritma Naive Bayes. *Jurnal Informatika*, 8(1), 884–898. <https://doi.org/10.1364/OFC.2009.OWD2>
- Bayati, M. A., & Jabbar, S. F. (2015). Developing a Spam Email Detector. *International Journal of Engineering and Innovative Technology (IJEIT)*, 5(2), 16–21.
- Bustami. (2014). Penerapan Algoritma Naive Bayes. *Jurnal Informatika*, 8(1), 884–898. <https://doi.org/10.1364/OFC.2009.OWD2>
- Dada, E. G., & Joseph, S. B. (2018). Random Forests Machine Learning Technique for Email Spam Filtering, 9(1), 29–36.
- Joseph, S. R., Hlomani, H., & Letsholo, K. (2016). Data Mining Algorithms : An Overview Academic Discipline And Sub-Disciplines Data Mining and Algorithms, 15, 6806–6813.
- Juang, D. (2016). Analisis spam dengan menggunakan naïve bayes. *Jurnal Teknovasi*, 3(1998), 51–57.
- Kalam, A. (2016). Advances in Electronics , Communication and Computing, (January). <https://doi.org/10.1007/978-981-10-4765-7>

Kruchten, P. (2003). *The Rational Unified Process : An Introduction*.

Mawarsih, A. (2014). PENGARUH ELECTRONIC MAIL SEBAGAI MEDIA KOMUNIKASI TERKADAP MENGERJAKAN TUGAS KULIAH MAHASISWA, 2(1), 334–348.

Rusland, N. F., Wahid, N., Kasim, S., & Hafit, H. (2017). Analysis of Naïve Bayes Algorithm for Email Spam Filtering across Multiple Datasets. *IOP Conference Series: Materials Science and Engineering*, 226(1). <https://doi.org/10.1088/1757-899X/226/1/012091>