

**ANALISIS PENERAPAN PENYARINGAN PESAN SPAM
PADA SMS MENGGUNAKAN METODE MULTINOMIAL
NAIVE BAYES**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

SOFYAN ADI SAPUTRA

19.11.3196

Kepada

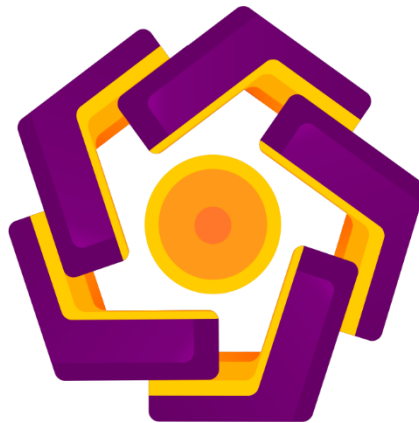
**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

**ANALISIS PENERAPAN PENYARINGAN PESAN SPAM
PADA SMS MENGGUNAKAN METODE MULTINOMIAL
NAIVE BAYES**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

SOFYAN ADI SAPUTRA

19.11.3196

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS PENERAPAN PENYARINGAN PESAN SPAM
PADA SMS MENGGUNAKAN METODE MULTINOMIAL
NAIVE BAYES**

yang disusun dan diajukan oleh

Sofyan Adi Saputra

19.11.3196

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 28 Februari 2023

Dosen Pembimbing,

Krisnawati, S.Si., M.T.

NIK. 190302038

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS PENERAPAN PENYARINGAN PESAN SPAM PADA SMS
MENGUNAKAN METODE MULTINOMIAL NAIVE BAYES**

yang disusun dan diajukan oleh

Sofyan Adi Saputra

19.11.3196

Telah dipertahankan di depan Dewan Penguji
pada tanggal 28 Februari 2023

Susunan Dewan Penguji

Nama Penguji

Eli Pujastuti, M.Kom
NIK. 190302227

Ika Nur Fajri, M.Kom
NIK. 190302268

Krisnawati, S.Si., M.T.
NIK. 190302038

Tanda Tangan




Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 28 Februari 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Sofyan Adi Saputra
NIM : 19.11.3196

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS PENERAPAN PENYARINGAN PESAN SPAM PADA SMS
MENGUNAKAN METODE MULTINOMIAL NAIVE BAYES**
Dosen Pembimbing : Krisnawati, S.Si., M.T.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 28 Februari 2023

Yang Menyatakan,



Sofyan Adi Saputra

HALAMAN PERSEMBAHAN

Alhamdulillahirobbil'alamiin, segala puji bagi Allah Subhanahu Wa Ta'ala yang telah mencurahkan rahmat dan karunia-Nya, sehingga Peneliti mampu menyelesaikan Skripsi dengan judul “**Analisis Penerapan Penyaringan Pesan Spam Pada Sms Menggunakan Metode Multinomial Naive Bayes**” ini dengan baik. Shalawat serta Salam tercurahkan kepada sang baginda Rasulullah Muhammad Shalallahu Alaihi Wassalam yang telah membimbing serta menunjukkan jalan dan petunjuk bagi umatnya. Semoga doa dan shalawat dapat menggiring kita semua menuju jalan yang lebih baik dan benar, baik di dunia maupun akhirat. Aamiin Ya Rabbal Alamin.

Karya ini Saya persembahkan untuk :

1. Ayahanda dan Ibunda tercinta, yang sampai saat ini masih terus memberikan doa, dukungan, serta motivasi
2. Seluruh keluarga saya yang telah memberikan segala dukungan, doa, serta semangat.
3. Kepada Ibu Krisnawati, S.Si., M.T. yang telah memberikan bimbingan dalam skripsi ini serta memberikan ilmu baru dalam pengerjaan skripsi.
4. Kepada teman-teman kelas 19 IF 10, terimakasih banyak sudah menemani serta membantu saya mulai dari awal perkuliahan hingga skripsi ini selesai disusun.
5. Kepada teman – teman kos ijo yang telah memberikan semangat dan menemani saya dalam pengerjaan skripsi.
6. Kepada Senior-senior yang telah membantu serta mengarahkan saya mulai dari sebelum masuk dunia perkuliahan hingga sekarang.
7. Almamater serta Kampus tercinta, Universitas AMIKOM Yogyakarta

KATA PENGANTAR



Assalamu 'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillah *allobbil'amin*, puji syukur kehadiran Allah Subhanahu Wa Ta'ala atas segala rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan Skripsi ini dalam rangka memenuhi sebagian persyaratan untuk memperoleh gelar Sarjana pada Program Studi Informatika Universitas Amikom Yogyakarta dengan judul “**Analisis Penerapan Penyaringan Pesan Spam Pada Sms Menggunakan Metode Multinomial Naive Bayes**”. Alhamdulillah skripsi ini dapat disusun dengan baik dan lancar atas bantuan, dukungan, serta bimbingan dari berbagai pihak. Untuk itu dalam kesempatan ini peneliti mengucapkan terima kasih kepada:

1. Ayahanda dan Ibunda tercinta, yang sampai saat ini masih terus memberikan doa, dukungan, serta motivasi.
2. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
3. Bapak Hanif Al Fatta, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta
4. Ibu Windha Mega Pradnya D, M.Kom, selaku Ketua Program Studi Informatika Universitas Amikom Yogyakarta.
5. Ibu Krisnawati, S.Si., M.T., Selaku dosen Pembimbing yang selalu memberikan bimbingan, waktu, dan arahan dalam pembuatan Skripsi ini.
6. Bapak Majid Rahardi, M.Eng, selaku Dosen Wali yang telah banyak membantu dalam membimbing selama masa perkuliahan.
7. Bapak/Ibu Dosen Penguji yang telah meluangkan waktu serta tenaga disaat proses pendadaran/sidang skripsi.
8. Seluruh Dosen Universitas Amikom yang telah memberikan ilmu selama perkuliahan.

9. Kerabat terdekat yang telah memberikan dukungan, doa, serta semangat dalam proses penyusunan skripsi.
10. Teman-teman seperjuangan yang telah menemani, memberikan semangat, serta motivasi untuk terus melangkah maju.
11. Semua pihak yang telah membantu dalam kelancaran penulisan skripsi ini baik langsung maupun tidak langsung yang tidak dapat peneliti sebutkan satu persatu.

Peneliti menyadari bahwa laporan ini masih jauh dari sempurna, meskipun demikian peneliti berharap semoga laporan ini bermanfaat bagi yang membacanya dan peneliti dengan senang hati menerima kritik dan saran yang membangun dari para pembaca.

Akhir kata peneliti berharap semoga hasil karya ini dapat berguna serta bermanfaat bagi perkembangan Teknologi dan Informasi pada khususnya, serta sebagai kajian bagi mahasiswa Universitas Amikom Yogyakarta lainnya dalam pengambilan Skripsi.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Yogyakarta, 28 Februari 2023

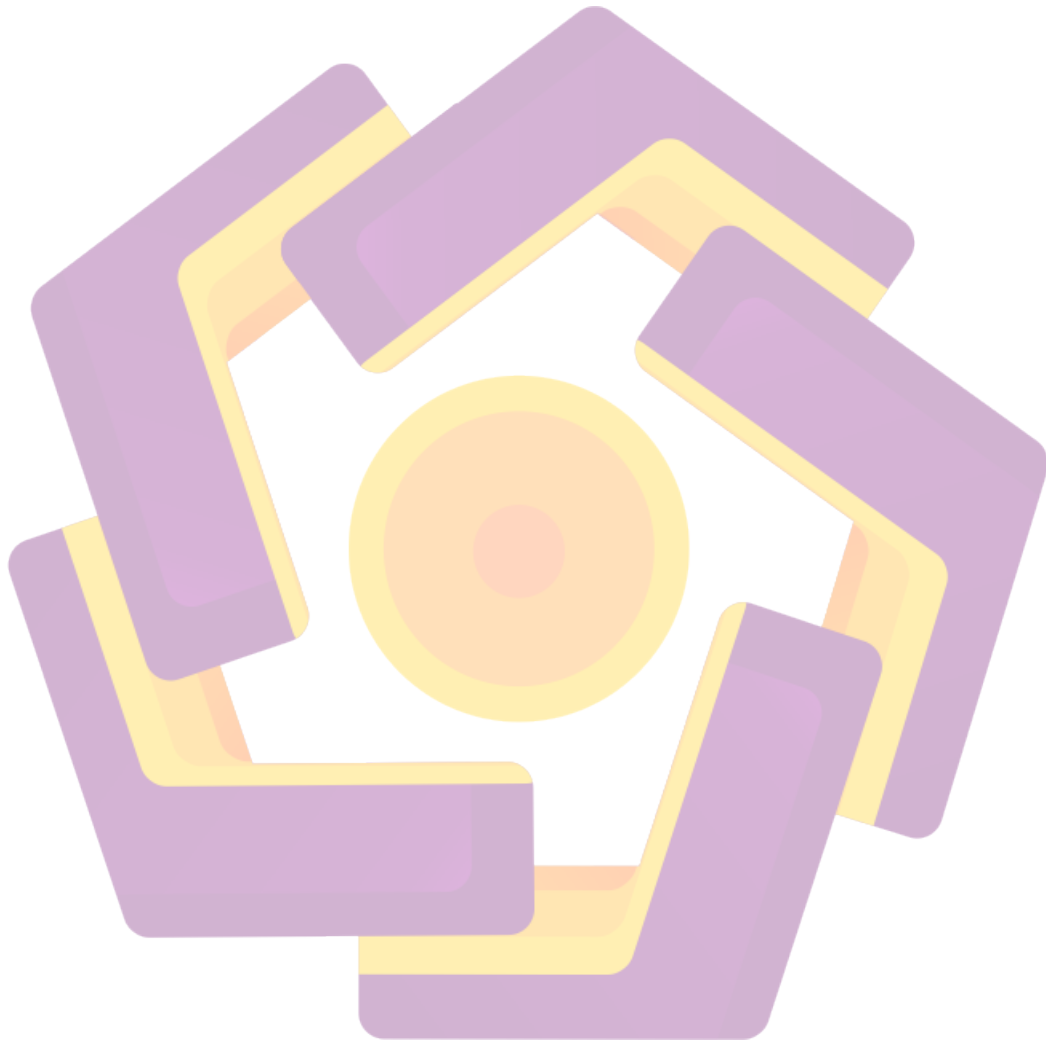
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iii
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
INTISARI	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	9
2.2.1 SMS.....	9
2.2.2 Spam	9
2.2.3 Klasifikasi	10
2.2.4 Text Preprocessing.....	11
2.2.5 Term Frequency Inverse Document Frequency (TF-IDF).....	13
2.2.6 Multinomial Naïve Bayes	13
2.3 Metode Pengujian	13
BAB III METODE PENELITIAN	16

3.1	Alur Penelitian	16
3.2	Alat dan Bahan.....	17
3.2.1	Alat Penelitian.....	17
3.2.2	Bahan Penelitian	17
3.3	Pengumpulan Data.....	18
3.4	Teks Preprocessing	20
3.4.1	Case Folding	20
3.4.2	Tokenizing	20
3.4.3	Stopword.....	21
3.4.4	Stemming	21
3.5	Pembobotan TF-IDF.....	22
3.6	Proses Klasifikasi.....	24
3.6.1	Multinomial Naïve Bayes	24
3.6.2	Pengklasifikasian Lanjutan	25
3.6.2.1	Bahasa Campuran.....	25
3.6.2.2	Istilah Gaul (<i>Slang</i>)	26
3.6.2.3	Simbol – Simbol Tertentu	26
3.7	Pengujian Model.....	28
BAB IV	HASIL DAN PEMBAHASAN	31
4.1	Analisis Model.....	31
4.1.1	Load Dataset	31
4.1.2	Visualisasi Data	32
4.1.3	Data Cleaning.....	36
4.1.4	Konversi Model Machine Learning	37
4.1.5	Implementasi Metode Naïve Bayes dan Pembobotan TF-IDF	43
4.1.6	Pengujian Confusion Matrix	44
4.2	Implementasi Model	45
4.2.1	Load Dataset	45
4.2.2	Preprocessing	46
4.2.3	Predict Function	50
4.2.4	Program Testing.....	51

4.3 Validasi Model.....	54
BAB V PENUTUP	56
5.1 Kesimpulan	56
5.2 Saran	56
REFERENSI	57



DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	7
Tabel 2.2. Contoh Tabel Confusion Matrix	14
Tabel 3.1. Confusion matrix	28
Tabel 4.1. Data Hasil Preprocessing	36
Tabel 4.2. Data Hasil Preprocessing Bahasa Indonesia	37
Tabel 4.3. Pembagian Data 20%	38
Tabel 4.4. Pembagian Data 20% dataset Bahasa Indonesia	38
Tabel 4.5. Pembagian Data 40%	39
Tabel 4.6. Pembagian Data 40% dataset Bahasa Indonesia	39
Tabel 4.7. Pembagian Data 60%	40
Tabel 4.8. Pembagian Data 60% dataset Bahasa Indonesia	40
Tabel 4.9. Pembagian Data 80%	41
Tabel 4.10. Pembagian Data 80% dataset Bahasa Indonesia	41
Tabel 4.11. Pembagian Data 99%	42
Tabel 4.12. Pembagian Data 99% dataset Bahasa Indonesia	42
Tabel 4.13 Hasil Perhitungan Confusion Matrix	44
Tabel 4.14. Pesan Teks dan Label	54
Tabel 4.15. Frekuensi Kata	54

DAFTAR GAMBAR

Gambar 3.1. Alur Penelitian	16
Gambar 3.2. Dataset SMS Spam Collection.....	18
Gambar 3.3. Penyajian Data Dalam Tabel.....	19
Gambar 3.4. Alur TD-IDF	23
Gambar 4.1. Import Library	31
Gambar 4.2. Load Dataset.....	31
Gambar 4.3. Load Dataset (Bahasa Indonesia).....	32
Gambar 4.4. Informasi Umum Dataset	32
Gambar 4.5. Diagram Perbandingan Jumlah Spam dan Ham	33
Gambar 4.6. Diagram Perbandingan Bahasa Indonesia.....	33
Gambar 4.7. Visualisasi Kepadatan Pesan Spam dan Ham	34
Gambar 4.8. Visualisasi Kepadatan Pesan Spam dan Ham Dataset Indonesia.....	34
Gambar 4.9. Informasi Frasa atau Kata pada Dataset.....	35
Gambar 4.10. Script Koversi Model	37
Gambar 4.11. Script matriks MNB	43
Gambar 4.12. Script matriks MNB (Bahasa Indonesia)	43
Gambar 4.13. Implementasi Load Dataset.....	45
Gambar 4.14. Proses Stopword.....	46
Gambar 4.15. Proses Stopword dengan Dataset Bahasa Indonesia	47
Gambar 4.16. Proses Case Folding dan Tokenizing	47
Gambar 4.17. Proses Case Folding dan Tokenizing Bahasa Indonesia	48
Gambar 4.18. Contoh Kata Spam dan Ham.....	48
Gambar 4.19. Kata Spam dan Ham Bahasa Indonesia	49
Gambar 4.20. Fungsi Prediksi.....	50
Gambar 4.21. Pengujian Spam.....	51
Gambar 4.22. Pengujian Ham Bahasa Indonesia	51
Gambar 4.23 Pengujian Ham	52
Gambar 4.24. Pengujian Ham Bahasa Indonesia	52
Gambar 4.25. Pengujian Lain	53

INTISARI

SMS (*Short Message Service*) merupakan salah satu layanan komunikasi yang perkembangannya paling pesat. Pesatnya perkembangan layanan SMS tidak hanya berdampak positif untuk kehidupan namun juga menimbulkan masalah yang serius dikarenakan oknum yang tidak bertanggungjawab. Pesan spam merupakan masalah umum yang berada pada berbagai layanan pengolahan dan pengiriman pesan khususnya layanan SMS. Pesan spam terkadang berisi berbagai macam pesan berupa promosi, iklan, dan sejenisnya. Namun tak jarang pesan spam juga berisi pesan sejenis tautan berbahaya yang berisi *virus*, *malware*, penipuan, dan sejenisnya.

Tujuan peneliti dalam melakukan penelitian yaitu untuk mengembangkan sistem yang secara otomatis dapat mengklasifikasikan dan memfilter pesan SMS sebagai *spam* atau *non-spam (ham)*. Sistem yang diusulkan menggunakan Algoritma Multinomial Naive Bayes untuk klasifikasi teks, yang merupakan metode dikembangkan dalam NLP (*Natural Language Processing*). Dataset yang digunakan untuk pelatihan (*training*) dan pengujian (*testing*) sistem adalah kumpulan pesan SMS berlabel *spam* atau *ham*.

Sistem filter ini dibuat dengan *Google Colab* dengan bahasa pemrograman Python. Hasil penelitian menunjukkan bahwa Algoritma Multinomial Naive Bayes adalah metode yang efektif untuk penyaringan teks SMS dengan nilai *accuracy*, *precision*, *recall*, dan *F1-Score* yang tinggi. Sistem mencapai akurasi 87% dalam mengklasifikasikan pesan SMS sebagai *spam* atau *ham* menggunakan dataset Bahasa Inggris dan 77% menggunakan dataset Bahasa Indonesia. Hasil penelitian ini menunjukkan keefektifan penggunaan metode Multinomial Naive Bayes dalam pemfilteran teks SMS dan memberikan dasar untuk penelitian selanjutnya pada bidang ini.

Kata kunci: SMS, Spam, Filter, Multinomial Naive Bayes

ABSTRACT

SMS (Short Message Service) is one of the fastest growing communication services. The rapid development of SMS services not only has a positive impact on life but also creates serious problems due to irresponsible persons. Spam messages are a common problem in various message processing and delivery services, especially SMS services. Spam messages sometimes contain various kinds of promotional messages, advertisements, and the like. But not infrequently spam messages also contain messages similar to malicious links that contain viruses, malware, fraud, and the like.

The aim of the researcher in conducting the research is to develop a system that can automatically classify and filter SMS messages as spam or non-spam (ham). The proposed system uses the Multinomial Naive Bayes Algorithm for text classification, which is a method developed in NLP (Natural Language Processing). The dataset used for system training and testing is a collection of SMS messages labeled spam or ham.

This filter system is made with Google Colab with the Python programming language. The results showed that the Naïve Bayes Multinomial Algorithm is an effective method for filtering SMS text with high accuracy, precision, recall, and F1-Score values. The system achieved 87% accuracy in classifying SMS messages as spam or ham using the English dataset and 77% using the Indonesian dataset. The results of this study demonstrate the effectiveness of using the Multinomial Naive Bayes method in filtering SMS text and provide a basis for further research in this field.

Keyword: *SMS, Spam, Filter, Multinomial Naïve Bayes*