

**EFEKTIVITAS DAN IMPLEMENTASI ALGORITMA NAIVE
BAYES UNTUK MENGANALISIS MENDETEKSI PESAN
PENIPUAN PADA SMS**

SKRIPSI

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



disusun oleh

MUHAMMAD GILANG DWI SAPUTRA

21.11.4233

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**EFEKTIVITAS DAN IMPLEMENTASI ALGORITMA NAIVE
BAYES UNTUK MENGANALISIS MENDETEKSI PESAN
PENIPUAN PADA SMS**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

MUHAMMAD GILANG DWI SAPUTRA
21.11.4233

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2024

HALAMAN PERSETUJUAN

SKRIPSI

**EFEKTIVITAS DAN IMPLEMENTASI ALGORITMA NAIVE BAYES
UNTUK MENGANALISIS MENDETEKSI PESAN PENIPUAN PADA SMS**

yang disusun dan diajukan oleh

Muhammad Gilang Dwi Saputra

21.11.4233

telah disetujui oleh Dosen Pembimbing Skripsi
pada 19 Desember 2024

Dosen Pembimbing,



Rumini, M.Kom

NIK. 190302246

HALAMAN PENGESAHAN

SKRIPSI

**EFEKTIVITAS DAN IMPLEMENTASI ALGORITMA NAIVE BAYES
UNTUK MENGANALISIS MENDETEKSI PESAN PENIPUAN PADA SMS**

yang disusun dan diajukan oleh

Muhammad Gilang Dwi Saputra

21.11.4233

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Desember 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Drs. Bambang Sudarvatno, M.M.,
NIK. 190302029

Dwi Nurani, S.Kom., M.Kom.,
NIK. 190302236

Rumini, S.Kom., M.Kom.,
NIK. 190302246



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Desember 2024

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Muhammad Gilang Dwi Saputra
NIM : 21.11.4233

Menyatakan bahwa Skripsi dengan judul berikut:

EFEKTIVITAS DAN IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MENGANALISIS MENDETEKSI PESAN PENIPUAN PADA SMS

Dosen Pembimbing : Rumini, M.Kom

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, 19 Desember 2024

Yang Menyatakan,



Muhammad Gilang Dwi Saputra

HALAMAN PERSEMBAHAN

Segala puji serta syukur saya panjatkan ke hadirat Allah SWT, atas limpahan rahmat, kasih sayang, dan karunia-Nya, sehingga saya dapat menyelesaikan skripsi ini yang berjudul "Efektivitas dan Implementasi Algoritma Naïve Bayes untuk Menganalisis dan Mendeteksi Pesan Penipuan Pada SMS". Karya ini adalah bukti dari usaha dan perjalanan yang penuh dengan tantangan selama masa perkuliahan. Dengan segala kerendahan hati, karya ini saya persembahkan kepada :

1. Saya sendiri, yang telah berjuang melewati berbagai rintangan selama masa studi. Saya bangga atas ketekunan dan semangat yang tidak pernah padam.
2. Kedua orang tua saya, yang selalu menjadi sumber inspirasi dan dukungan tanpa batas. Terima kasih atas doa, cinta, dan pengorbanan yang telah mereka berikan.
3. Keluarga dan saudara-saudara, Kakak Nanda Eka Putri Anna Pangestu, S.Gz, dan Adik Asyfa Tri Puspita Sari, yang selalu memberikan dukungan moral dan motivasi.
4. Ibu Rumini, M.Kom, dosen pembimbing saya, yang telah membimbing dan memberikan arahan selama penulisan skripsi ini. Terima kasih atas kesabaran dan bimbingannya.
5. Perempuan yang saya sayang, Sitti Mutiara Nur Aliyfa, yang selalu memberikan dukungan dan semangat tanpa henti setiap harinya.
6. Teman-teman dan sahabat, yang selalu menemani dan memberikan semangat selama masa kuliah hingga penyusunan skripsi ini. Terima kasih atas canda tawa dan kebersamaannya.
7. Almamater tercinta, tempat saya menimba ilmu dan berkembang. Semoga karya ini dapat berkontribusi bagi kemajuan ilmu pengetahuan di masa depan.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT, atas limpahan rahmat, hidayah, serta karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "Efektivitas dan Implementasi Algoritma Naïve Bayes untuk Menganalisis dan Mendeteksi Pesan Penipuan Pada SMS". Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan studi di Program Sarjana Universitas Amikom Yogyakarta. Proses penulisan skripsi ini tentu tidak lepas dari bantuan, dukungan, dan bimbingan banyak pihak. Oleh karena itu, penulis ingin mengucapkan rasa terima kasih yang sebesar-besarnya kepada :

1. Bapak Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom, Selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Windha Mega Pradnya D, M.Kom selaku ketua program studi SI Informatika.
4. Ibu Ruminí, M.Kom, selaku dosen pembimbing saya yang telah banyak memberikan pengarahán bagi penulis dalam pembuatan skripsi.
5. Bapak-Ibu Dosen Universitas Amikom Yogyakarta, atas ilmu dan pengetahuan yang telah diberikan kepada penulis selama duduk di bangku kuliah.

Penulis tentunya menyadari bahwa pembuatan skripsi ini masih banyak sekali kekurangan dan kelemahannya. Oleh karena itu penulis berharap kepada semua pihak agar dapat menyampaikan kritik dan saran yang membangun untuk menambah kesempurnaan skripsi ini. Dengan rasa syukur dan harapan, semoga skripsi ini dapat memberikan kontribusi positif bagi pengembangan ilmu pengetahuan dan teknologi, serta bermanfaat bagi mereka yang mempelajari dan menggunakannya.

Yogyakarta, 19 Desember 2024

Muhammad Gilang Dwi Saputra

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
INTISARI	xiv
ABSTRACT	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	14
2.2.1 Machine Learning	14

2.2.2	<i>Data Mining</i>	14
2.2.3	<i>Naive Bayes</i>	16
2.2.4	<i>Confusion Matrix</i>	17
2.2.5	<i>Accuracy</i>	18
2.2.6	<i>Precision (Positive Predictive Value)</i>	18
2.2.7	<i>Recall atau Sensitivity (True Positive Rate)</i>	19
2.2.8	Pembagian Dataset	19
2.2.9	SMS (<i>Short Message Service</i>)	19
2.2.10	<i>Google Colab</i>	20
2.2.11	<i>GitHub</i>	21
2.2.12	<i>Streamlit</i>	21
BAB III METODE PENELITIAN		22
3.1	Objek Penelitian	22
3.2	Alur Penelitian	22
3.2.1	Pengumpulan Data	24
3.2.2	Pelabelan	24
3.2.3	<i>Preprocessing</i> Data	25
3.2.4	Visualisasi Data	27
3.2.5	Transformation (<i>TF-IDF</i>) & (<i>N-Gram</i>)	28
3.2.6	Pembagian Data	28
3.2.7	Klasifikasi Data (Pembuatan Model)	29
3.2.8	Uji Model Dan Evaluasi Model (<i>Confusion Matrix</i>)	30
3.2.9	<i>Deployment</i> Aplikasi	31
3.3	Alat dan Bahan	32
3.3.1	Data Penelitian	32

3.3.2	Alat/instrumen.....	32
BAB IV HASIL DAN PEMBAHASAN		34
4.1	Pengumpulan Data	34
4.2	Pelabelan Data	35
4.3	<i>Preprocessing</i> Data	37
4.2.1	<i>Case Folding</i>	37
4.2.2	<i>Word Normalization</i>	39
4.2.3	<i>StopWord Removal</i>	40
4.2.4	<i>Stemming</i>	41
4.4	Visualisasi Data	42
4.5	Transformasi <i>TF-IDF</i> dan <i>N-Gram</i>	45
4.6	Pembagian Data	46
4.7	Klasifikasi Data (Pembuatan Model).....	47
4.8	Uji Model Dan Evaluasi Model	48
4.9	<i>Deployment</i> Aplikasi.....	51
BAB V PENUTUP		65
5.1	Kesimpulan	65
5.2	Saran	65
REFERENSI		67
LAMPIRAN.....		70

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	9
Tabel 3. 1 Pelabelan	24
Tabel 3. 2 Case Folding	25
Tabel 3. 3 Word Normalization	26
Tabel 3. 4 Stopword Removal	26
Tabel 3. 5 Stemming	27
Tabel 3. 6 Word Cloud	27
Tabel 3. 7 TF-IDF & N-Gram	28
Tabel 3. 8 Pembagian Data	29
Tabel 3. 9 Klasifikasi Data (Pembuatan Model)	29
Tabel 3. 10 Representasi Confusion Matrix 3x3	31
Tabel 3. 11 Deployment Aplikasi	31

DAFTAR GAMBAR

Gambar 2. 1 Knowledge Discovery pada Database	15
Gambar 2. 2 Confusion Matrix	17
Gambar 3. 1 Alur Penelitian	23
Gambar 4. 1 Proses Memasukan data	34
Gambar 4. 2 Melihat Data	34
Gambar 4. 3 Rekapitulasi Data	35
Gambar 4. 4 Proses Pelabelan Data	35
Gambar 4. 5 Hasil Pelabelan Data	36
Gambar 4. 6 Distribusi Data Setelah Pelabelan Data	36
Gambar 4. 7 Distribusi Data Dalam Persen	37
Gambar 4. 8 Proses Case Folding	38
Gambar 4. 9 Proses Case Folding	38
Gambar 4. 10 Sebelum Case Folding	38
Gambar 4. 11 Sebelum Case Folding	38
Gambar 4. 12 Sebelum Case Folding	38
Gambar 4. 13 Sesudah Case Folding	38
Gambar 4. 14 Sesudah Case Folding	38
Gambar 4. 15 Proses Word Normalization	39
Gambar 4. 16 Sebelum Word Normalization	39
Gambar 4. 17 Sebelum Word Normalization	39
Gambar 4. 18 Sesudah Word Normalization	40
Gambar 4. 19 Sesudah Word Normalization	40
Gambar 4. 20 Proses StopWord Removal	40
Gambar 4. 21 Sebelum Stopword Removal	41
Gambar 4. 22 Sesudah Stopword Removal	41
Gambar 4. 23 Proses Stemming	41
Gambar 4. 24 Sebelum Stemming	42
Gambar 4. 25 Sesudah Stemming	42
Gambar 4. 26 Perbandingan Setelah Dilakukan Preprocessing	42

Gambar 4. 27 Wordcloud Sms Normal	43
Gambar 4. 28 Most Common Words SMS Normal	43
Gambar 4. 29 Wordcloud Sms Penipuan	44
Gambar 4. 30 Most Common Words SMS Penipuan	44
Gambar 4. 31 Wordcloud Sms Promo	45
Gambar 4. 32 Most Common Words SMS Promo	45
Gambar 4. 33 Proses Pembobotan	46
Gambar 4. 34 Hasil Pembobotan	46
Gambar 4. 35 Proses Pembagian Data	47
Gambar 4. 36 Menampilkan Pembagian Data	47
Gambar 4. 37 Hasil Pembagian Data	47
Gambar 4. 38 Proses Klasifikasi Algoritma Naïve Bayes	48
Gambar 4. 39 Proses Input Manual Prediksi	48
Gambar 4. 40 Output Input Manual	49
Gambar 4. 41 HeatMap Confusion Matrix	50
Gambar 4. 42 Classification Report Naïve Bayes	50
Gambar 4. 43 Proses Menyimpan Model	51
Gambar 4. 44 Folder Kumpulan Model Dan Data	51
Gambar 4. 45 Anaconda Installers	52
Gambar 4. 46 Visual Studio Code Installers	52
Gambar 4. 47 Anaconda Prompt	53
Gambar 4. 48 ENV Setting	53
Gambar 4. 49 Pengecekan Python Dan Conda	53
Gambar 4. 50 Instalasi Python Language	53
Gambar 4. 51 Proses Install Streamlit	54
Gambar 4. 52 Pengecekan Streamlit	54
Gambar 4. 53 Proses Deployment	55
Gambar 4. 54 Proses Deployment	56
Gambar 4. 55 Proses Load Pada Aplikasi	57
Gambar 4. 56 Proses Deployment	57

Gambar 4. 57 Proses Sign In Github	58
Gambar 4. 58 Pembuatan Repository Pada Github	58
Gambar 4. 59 Proses Mengunggah File	59
Gambar 4. 60 Mengunggah File Pada Repository	59
Gambar 4. 61 Proses Sign In Streamlit	60
Gambar 4. 62 Proses Create Aplikasi	60
Gambar 4. 63 Proses Deployment	61
Gambar 4. 64 Proses Deployment	61
Gambar 4. 65 Proses Hosting Deployment	62
Gambar 4. 66 Tampilan Aplikasi Setelah Hosting dan Deployment Berhasil	62
Gambar 4. 67 Uji Coba Aplikasi	63
Gambar 4. 68 Uji Coba Aplikasi	63
Gambar 4. 69 Uji Coba Aplikasi	64



INTISARI

Penelitian ini mengeksplorasi deteksi penipuan pada pesan singkat (SMS) menggunakan algoritma *Naive Bayes*. Meskipun SMS kurang populer dibandingkan aplikasi chat lainnya, SMS tetap relevan terutama dalam transaksi online dan *e-commerce*. Namun, SMS sering disalahgunakan untuk melakukan penipuan. Penelitian ini bertujuan untuk mengembangkan sistem deteksi SMS penipuan dengan menggunakan algoritma *Naive Bayes* yang dikenal sederhana namun efektif dalam klasifikasi teks. Dengan menggunakan dataset 1143 SMS berbahasa Indonesia yang telah dibersihkan, dinormalisasi, dan diberi label. Representasi teks menggunakan *TF-IDF* dan *N-Gram*. Model *Naive Bayes* mencapai akurasi 91%, dengan *precision*, *recall*, dan *F1-Score* masing-masing 89%, 90%, dan 89%. Pola kata dalam SMS normal, penipuan, dan promo diidentifikasi dan divisualisasikan dengan word cloud. Meskipun ada kelemahan dalam asumsi independensi fitur, *Naive Bayes* terbukti efektif untuk deteksi SMS penipuan. Penelitian selanjutnya disarankan untuk mengeksplorasi algoritma lain seperti *Random Forest*, *SVM*, dan *Neural Networks*, serta analisis metadata dan pola perilaku pengguna untuk meningkatkan deteksi. Penelitian ini memberikan kontribusi dalam pengembangan sistem deteksi penipuan SMS yang lebih aman dan efisien.

Kata kunci: SMS, *Naive Bayes*, Klasifikasi Teks, *TF-IDF*, *N-Gram*.

ABSTRACT

This research explores the detection of fraud in short messages (SMS) using the Naive Bayes algorithm. Although SMS is less popular compared to other chat applications, it remains relevant, especially in online transactions and e-commerce. However, SMS is often misused for fraudulent purposes. This study aims to develop an SMS fraud detection system using the Naive Bayes algorithm, which is known for being simple yet effective in text classification. The study uses a dataset of 1143 Indonesian SMS that have been cleaned, normalized, and labeled. Text representation utilizes TF-IDF and N-Gram. The Naive Bayes model achieved an accuracy of 91%, with precision, recall, and F1-Score of 89%, 90%, and 89%, respectively. Word patterns in normal, fraudulent, and promotional SMS were identified and visualized using word clouds. Despite the limitation of feature independence assumptions, Naive Bayes has proven effective for SMS fraud detection. Future research is suggested to explore other algorithms such as Random Forest, SVM, and Neural Networks, as well as metadata analysis and user behavior patterns to enhance detection. This study contributes to the development of a safer and more efficient SMS fraud detection system.

Keyword: SMS, Naive Bayes, Klasifikasi Teks, TF-IDF, N-Gram