

**ANALISIS ALGORITMA NAIVE BAYES UNTUK PREDIKSI
RISIKO KANKER PARU-PARU BERDASARKAN
FAKTOR GAYA HIDUP**

SKRIPSI NON-REGULER JALUR SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika



Disusun oleh :

SHEILA ANGGUN VABILLA

22.11.4578

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS ALGORITMA NAIVE BAYES UNTUK PREDIKSI
RISIKO KANKER PARU-PARU BERDASARKAN
FAKTOR GAYA HIDUP**

SKRIPSI NON-REGULER JALUR SCIENTIST

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



Disusun oleh :

SHEILA ANGGUN VABILLA

22.11.4578

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI NON-REGULER JALUR SCIENTIST

**ANALISIS ALGORITMA NAIVE BAYES UNTUK PREDIKSI
RISIKO KANKER PARU-PARU BERDASARKAN
FAKTOR GAYA HIDUP**

yang disusun dan diajukan oleh

Sheila Anggun Vabilla

22.11.4578

telah disetujui oleh Dosen Pembimbing
pada tanggal 20 Januari 2026

Dosen Pembimbing,



Maito Rahardi, S.Kom., M.Eng
NIK. 190302393

HALAMAN PENGESAHAN

SKRIPSI NON-REGULER JALUR SCIENTIST

ANALISIS ALGORITMA NAIVE BAYES UNTUK PREDIKSI
RISIKO KANKER PARU-PARU BERDASARKAN
FAKTOR GAYA HIDUP

yang disusun dan diajukan oleh

Sheila Anggun Vabilla
22.11.4578

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Januari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Eli Pujastuti S.Kom., M.Kom.
NIK. 190302227

Subektiningsih S.Kom., M.Kom.
NIK. 190302413

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393



Laporan ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Sheila Anggun Vabilla

NIM : 22.11.4578

Menyatakan bahwa Laporan dengan judul berikut:

Analisis Algoritma Naïve Bayes Untuk Prediksi Risiko Kanker Paru-Paru Berdasarkan Faktor Gaya Hidup

Dosen Pembimbing : Majid Rahardi, S.Kom., M.Eng

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 kegiatan 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 ketidak-benaran 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, 20 Januari 2026

Yang Menyatakan,



Sheila Anggun Vabilla

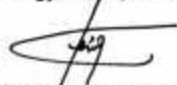
KATA PENGANTAR

Segala puji dan ucapan rasa syukur panjatkan kepada Allah SWT, atas berkat rahmat serta hidayah-Nya sehingga penulis dapat menyelesaikan laporan sidang ini dengan baik. Penyusunan laporan ini merupakan bagian dari proses penelitian sekaligus menjadi pemenuhan tanggung jawab akademik. Penyusunan laporan sidang ini terlaksana berkat arahan, bimbingan, doa dan dukungan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Orang tua dan kakak-kakak tercinta, yang senantiasa memberikan doa, dukungan, serta kasih sayang yang tiada henti,
2. Prof. Dr. Mohammad Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
3. Prof. Dr. Kusriani, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Eli Pujastuti, M.Kom selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
5. Majid Rahardi, S.Kom., M.Eng selaku Dosen Pembimbing yang memberikan arahan serta bimbingan sehingga penelitian dan penyusunan laporan ini dapat diselesaikan dengan baik.
6. Tim Dosen Penguji, atas saran, kritik, dan masukan yang bersifat membangun, sehingga membantu penulis dalam menyempurnakan laporan penelitian ini.

Penulis menyadari laporan ini masih memiliki keterbatasan, sehingga saran dan kritik yang membangun sangat diharapkan. Semoga laporan ini bermanfaat bagi pembaca dan pihak yang berkepentingan.

Yogyakarta, 3 Januari 2026



Sheila Anggun Vabilla

DAFTAR ISI

Halaman Judul.....	i
Halaman Persetujuan.....	ii
Halaman Pengesahan	iii
Halaman Pernyataan Keaslian Karya.....	iv
Kata Pengantar.....	v
Daftar Isi.....	vi
Daftar Gambar.....	viii
Daftar Lampiran.....	ix
Daftar Lambang dan Singkatan.....	x
Daftar Istilah.....	xi
Intisari	xii
<i>Abstract</i>	xiii
Bab I Pendahuluan	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah.....	2
1.4. Tujuan.....	3
Bab II Tinjauan Pustaka	4
2.1. Studi Literatur	4
2.2. Landasan Teori.....	5
BAB III Metode Penelitian	10
3.1. Jenis Penelitian.....	10
3.2. Objek Penelitian.....	10
3.3. Alur Penelitian	11
3.4. Sumber Data.....	12
3.5. Pre-processing Data	13
3.6. Split Data.....	15
3.7. Penyeimbangan data menggunakan SMOTE	16

3.8.	Seleksi Fitur Menggunakan Metode Mutual Information.....	17
3.9.	Evaluasi Model.....	18
3.10.	Algoritma Gaussian Naive Bayes	18
BAB IV Pembahasan		20
4.1.	Pre-Proseccing.....	20
4.2.	Split Data.....	22
4.3.	Normalisasi Data.....	22
4.4.	SMOTE	23
4.5.	Seleksi Fitur dengan Metode Mutual Information	23
4.6.	Evaluasi Model.....	25
BAB V Kesimpulan		27
5.1.	Kesimpulan	27
5.2.	Saran.....	28
Referensi		29
LAMPIRAN.....		33
Curriculum Vitae		33
Lampiran dan Bukti Pendukung.....		34
a.	Letter of Acceptance (LOA).....	34
b.	Lembar Review	35
c.	Bukti Terbit	39
d.	Bukti pembayaran	40

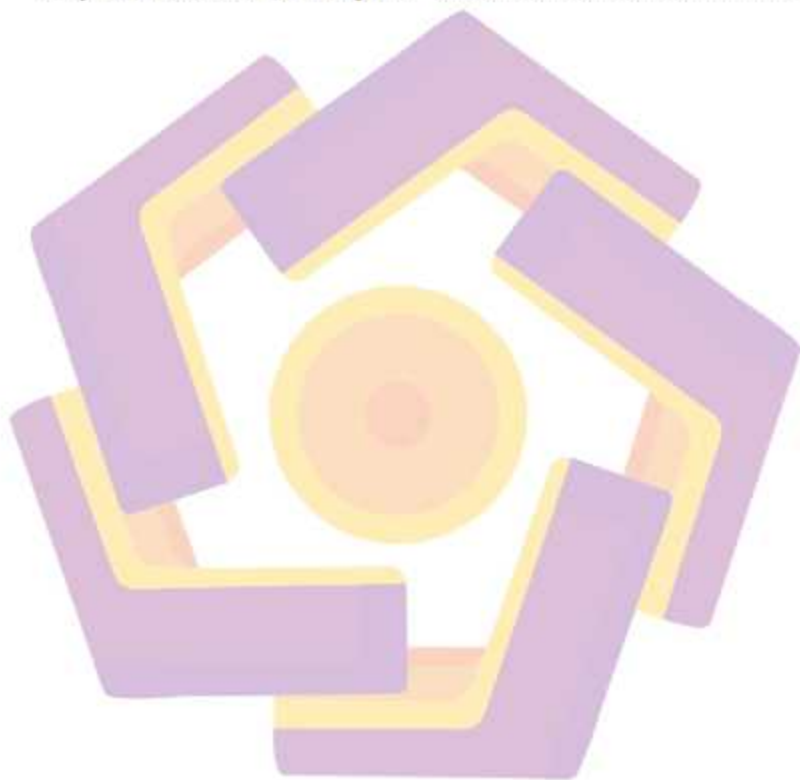
DAFTAR GAMBAR

Gambar 3.3.1 Alur Penelitian	11
Gambar 4.1.1.1 Hasil Missing Value	20
Gambar 4.1.1.2 Hasil Penyesuaian	21
Gambar 4.2.1 Hasil Split Data	22
Gambar 4.3.1 Hasil Normalisasi Data	23
Gambar 4.4.1 Hasil sesudah SMOTE	23
Gambar 4.5.1 Hasil Mutual Information	24
Gambar 4.6.1 Hasil Classification Report	25
Gambar 4.6.2 Hasil Confusion Matrix	26



DAFTAR LAMPIRAN

Lampiran 1. Curriculum Vitae	33
Lampiran 2. LOA	34
Lampiran 3. Lembar Bukti Riview	35
Lampiran 4. Bukti jurnal telah terbit	39
Lampiran 5. Bukti Terindex	39
Lampiran 6. Bukti submission fee jurnal	40



DAFTAR LAMBANG DAN SINGKATAN



GNB	<i>Gaussian Naive Bayes</i>
NB	<i>Naive Bayes</i>
SMOTE	<i>Synthetic Minority Over-sampling Technique</i>
MI	<i>Mutual Information</i>
SCLC	<i>Small Cell Lung Cancer</i>
NSCLC	<i>Non-Small Cell Lung Cancer</i>
CV	<i>Cross Validation</i>
z	Nilai hasil normalisasi
x	Nilai asli dari suatu fitur
μ	Nilai rata-rata dari fitur
σ	Simpangan baku fitur
x	Nilai fitur
μ_y	Nilai rata-rat fitur pada kelas y
σ_y	Simpangan baku fitur pada kelas y .

DAFTAR ISTILAH

Kanker Paru-paru	Penyakit akibat pertumbuhan sel abnormal pada jaringan paru-paru yang tidak terkendali.
Risiko Kanker	Tingkat kemungkinan seseorang mengalami kanker paru-paru berdasarkan faktor gaya hidup.
<i>Machine Learning</i>	Cabang kecerdasan buatan yang memungkinkan sistem belajar dari data tanpa pemrograman eksplisit.
Klasifikasi	Proses pengelompokan data ke dalam kelas tertentu.
<i>Pre-processing</i>	Tahap awal pengolahan data untuk meningkatkan kualitas data sebelum pemodelan.
Normalisasi	Proses penyetaraan skala data agar setiap fitur memiliki kontribusi yang seimbang.
<i>Confusion Matrix</i>	Matriks evaluasi hasil klasifikasi model.
<i>SMOTE</i>	Teknik oversampling untuk menyeimbangkan data dengan menambahkan sampel sintetis pada kelas minoritas.
<i>Mutual Information</i>	Ukuran statistik untuk menilai tingkat keterkaitan antara fitur dan variabel target.
<i>Classification Report</i>	Ringkasan hasil evaluasi model klasifikasi yang menampilkan nilai precision, recall, f1-score, dan akurasi.
<i>Heatmap</i>	Visualisasi data berbentuk peta warna yang menunjukkan tingkat hubungan atau nilai antar variabel.
<i>Overfitting</i>	Kondisi ketika model terlalu menyesuaikan diri dengan data latih sehingga kinerjanya menurun pada data baru.
<i>Encoding</i>	Proses mengubah data kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma pembelajaran mesin.
<i>Standard deviation</i>	Ukuran statistik yang menunjukkan tingkat sebaran data terhadap nilai rata-rata.
Akurasi	Persentase prediksi model yang benar dari seluruh data.

INTISARI

Kanker paru-paru merupakan salah satu jenis kanker dengan tingkat kematian tertinggi di dunia, yang seringkali sulit dideteksi pada stadium awal karena gejalanya minimal. Studi ini bertujuan untuk membangun model prediksi risiko kanker paru-paru berdasarkan faktor gaya hidup menggunakan algoritma *Gaussian Naive Bayes*. Keseimbangan data diatasi menggunakan *Synthetic Minority Over-sampling Technique (SMOTE)*, dan pemilihan fitur dilakukan menggunakan *Mutual Information*. Dataset yang digunakan terdiri dari 1000 data pasien dengan 26 fitur yang berkaitan dengan gaya hidup dan faktor lingkungan. Validasi model dilakukan menggunakan *5-fold Stratified Cross Validation*, dan dievaluasi berdasarkan akurasi, presisi, recall, *f1-score* dan matriks. Hasil menunjukkan bahwa penerapan SMOTE berhasil meningkatkan akurasi model hingga 91,00% dengan rata-rata precision sebesar 92,00%, recall sebesar 91,00%, dan *f1-score* sebesar 91,00%. Fitur "Perokok Pasif" dan "Batuk Berdarah" diidentifikasi sebagai faktor yang paling berpengaruh dalam prediksi. Hasil penelitian ini menunjukkan bahwa kombinasi *Gaussian Naive Bayes* dengan SMOTE dan *Mutual Information* mampu menghasilkan model prediksi yang akurat.

Kata kunci: kanker paru-paru, gaya hidup, Gaussian Naive Bayes, SMOTE, model Mutual Information

ABSTRACT

Lung cancer is one of the types of cancer with the highest mortality rate in the world, which is often difficult to detect in the early stages due to minimal symptoms. This study aims to build a lung cancer risk prediction model based on lifestyle factors using the Gaussian Naive Bayes algorithm. Data fit is addressed using the Synthetic Minority Over-sampling Technique (SMOTE), and feature selection is carried out using the Mutual Information. The dataset used consists of 1000 patient data with 26 features related to lifestyle and environmental factors. Model validation is carried out using 5-fold Stratified Cross Validation, and evaluated based on accuracy, precision, recall, and confusion matrices. The results show that the application of SMOTE successfully increases the model accuracy to 91.00% with an average precision of 92.00%, recall of 91.00%, and f1-score of 91.00%. The features "Passive Smoker" and "Coughing up Blood" are identified as the most influential factors in the prediction. The results of this study indicate that the combination of Gaussian Naive Bayes with SMOTE and Mutual Information is able to produce an accurate prediction model.

Keyword: lung cancer, lifestyle, Gaussian Naive Bayes, SMOTE, model Mutual Information