



**ANALISIS PREDIKSI PENYAKIT KANKER PARU-PARU  
MENGUNAKAN NAÏVE BAYES**

**SKRIPSI**

**RAFLI RAMADHAN**

**1810512047**

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**2022**



**ANALISIS PREDIKSI PENYAKIT KANKER PARU-PARU  
MENGUNAKAN NAÏVE BAYES**

**TUGAS AKHIR**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar  
Sarjana Ilmu Komputer**

**RAFLI RAMADHAN**

**1810512047**

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**2022**

# PERNYATAAN ORISINILITAS

## PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Rafli Ramadhan

NIM : 1810512047

Fakultas : Ilmu Komputer

Program Studi : S1 – Sistem Informasi

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku:

Jakarta, 21 Juli 2022  
Yang Menyatakan,



Rafli Ramadhan

## **PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai civitas akademis Universitas Pembangunan Nasional Veteran Jakarta, saya bertanda tangan di bawah ini:

Nama : Rafli Ramadhan

NIM : 1810512047

Fakultas : Ilmu Komputer

Program Studi : Sistem Informasi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan pelayanan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right ) atas karya ilmiah saya yang berjudul :

### **Analisis Prediksi Penyakit Kanker Paru-Paru Menggunakan Naïve Bayes**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Tugas Akhir saya selama mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada Tanggal : 22 Agustus 2022

Yang Menyatakan



(Rafli Ramadhan)

## LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir berikut :

Nama : Rafli Ramadhan  
NIM : 1810512047  
Program Studi : SI Sistem Informasi  
Judul : ANALISIS PREDIKSI PENYAKIT KANKER PARU-PARU  
MENGUNAKAN NAIVE BAYES

Telah disetujui untuk diujikan oleh Tim Penguji pada Sidang Tugas Akhir sebagai bagian persyaratan yang diperlukan untuk mengikuti Sidang Tugas Akhir pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

Ati Zaidiah, S.Kom., MTI  
Penguji I

Sarika M. Kom  
Penguji II

Dr. Ermatita M. Kom  
Pembimbing 1

I Wawan Widi P., S.Kom., MTI.  
Pembimbing 2



Dr. Ermatita M. Kom.  
Dekan Fakultas Ilmu Komputer

Helena Nurramdhani Irmanda, S.Pd., M.Kom.  
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Persetujuan : 21 Juli 2022



## KATA PENGANTAR

Puji dan syukur atas kehadiran Allah SWT yang telah memberikan rahmat dan hidayah-Nya kepada kita semua sehingga tugas akhir ini dapat terselesaikan dan topik yang dipilih adalah **SISTEM PREDIKSI PENYAKIT KANKER PARU-PARU MENGGUNAKAN *NAÏVE BAYES*** . Dalam kesempatan ini, penulis ingin mengucapkan ucapan terima kasih yang sebesar-besarnya atas bimbingan dan kerja sama yang diberikan kepada penulis, dan untuk itu penulis ucapkan terima kasih kepada :

1. Allah SWT, yang telah memberikan rahmat serta hidayah-Nya.
2. Keluarga Kakak dan Ibu dan keluarga terdekat saya.
3. Dr. Ermatita, M.Kom selaku Dekan Fakultas Ilmu Komputer dan selaku dosen pembimbing saya.
4. I Wayan Widi P.,S.Kom., MTI selaku dosen pembimbing saya.
5. Helena Nurramdhani Irmanda S.Pd., sebagai Ketua Program Studi Sistem Informasi.
6. Erik, Rafa, Anggi, dan semua teman RAID yang memberikan saya semangat dalam mengerjakan tugas akhir ini.
7. Teman seangkatan 2018 yang tidak bisa saya sebutkan.
8. Seluruh pihak yang terlibat dalam penyusunan Tugas Akhir ini yang tidak bisa saya sebutkan.

Masih banyak kekurangan dalam penyusunan tugas akhir ini, sehingga selaku penulis sangat membutuhkan kritik dan saran dari saudara semua agar bisa lebih baik kedepannya.

Jakarta, 27 Juni 2022

Penulis

## DAFTAR ISI

PERNYATAAN ORISINILITAS.....	ii
LEMBAR PENGESAHAN.....	iii
ABSTRAK.....	Error! Bookmark not defined.
KATA PENGANTAR.....	v
DAFTAR ISI.....	vi
DAFTAR LAMBANG.....	xi
DAFTAR LAMPIRAN.....	xii
BAB I.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	2
1.3. Tujuan Penelitian.....	3
1.4. Manfaat Penelitian.....	3
1.5. Ruang Lingkup.....	3
1.6. Sistematika Penulisan.....	3
BAB II.....	5
2.1. Kanker.....	5
2.2. Kanker Paru-paru.....	6
2.3. Analisis.....	7
2.4. Python.....	7
2.5. <i>Data Mining</i> .....	8
2.6. CRISP-DM.....	11
2.7. Naïve Bayes.....	13

2.8. <i>Confusion Matrix</i> .....	16
2.9. Studi Literatur .....	17
<b>BAB III</b> .....	<b>20</b>
3.1. Jenis Penelitian .....	20
3.2. Studi Awal .....	21
3.3. Pemahaman Data .....	21
3.4. Pengumpulan Data .....	21
3.5. <i>Data Mining</i> .....	22
3.6. Uji Coba Model .....	22
3.7. Evaluasi Hasil .....	22
3.8. Alat Bantu Penelitian .....	23
3.9. Jadwal Penelitian .....	23
<b>BAB IV</b> .....	<b>25</b>
4.1. Pengumpulan Data .....	25
4.2. Data Preprocessing .....	28
4.2.1. Data Cleaning (Pembersihan Data) .....	30
4.2.2. Data Selection (Seleksi Data) .....	30
4.2.3. Data Transformation (Transformasi Data) .....	31
4.3. Algoritma <i>Naïve Bayes</i> .....	32
4.4. Perbandingan Dengan Menggunakan Python .....	55
4.5. Evaluasi .....	58
4.6. Implementasi Sistem .....	60
4.6.1. Halaman Home .....	61
4.6.3. Halaman Input Data .....	62



4.6.4. Blackbox Testing.....	63
<b>BAB V .....</b>	<b>64</b>
<b>KESIMPULAN .....</b>	<b>64</b>
<b>5.1. Kesimpulan .....</b>	<b>64</b>
<b>5.2. Saran .....</b>	<b>64</b>
<b>DAFTAR PUSTAKA .....</b>	<b>65</b>
<b>LAMPIRAN .....</b>	<b>69</b>

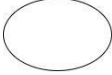


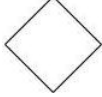
## DAFTAR TABEL

Tabel 3.1. Perkiraan Jadwal Penelitian .....	24
Tabel 4.1. Ringkasan Variabel <i>Dataset</i> .....	25
Tabel 4.2. Klasifikasi <i>Data Training</i> dan <i>Testing</i> .....	33
Tabel 4.3. Spesifikasi <i>Data Training</i> .....	33
Tabel 4.4. <i>Conditional Probability Table Age</i> .....	36
Tabel 4.5. <i>Conditional Probability Table Sex</i> .....	37
Tabel 4.6. <i>Conditional Probability Table Smoking Type</i> .....	38
Tabel 4.7. <i>Conditional Probability Table Yellow Fingers Type</i> .....	40
Tabel 4.8. <i>Conditional Probability Table Chronic Disease Type</i> .....	41
Tabel 4.9. <i>Conditional Probability Table Fatigue Type</i> .....	42
Tabel 4.10. <i>Conditional Probability Table Wheezing Type</i> .....	43
Tabel 4.11. <i>Conditional Probability Table Coughing Type</i> .....	45
Tabel 4.12. <i>Conditional Probability Table Shortness Of Breath Type</i> .....	46
Tabel 4.13. <i>Conditional Probability Table Swallowing Difficult Type</i> .....	47
Tabel 4.14. <i>Conditional Probability Table Chest Pain Type</i> .....	48
Tabel 4.15. Contoh Data Uji .....	49
Tabel 4.16. <i>Overview YES</i> .....	51
Tabel 4.17. <i>Overview NO</i> .....	53
Tabel 4.18. <i>Confusion Matrix Percobaan Pertama</i> .....	59
Tabel 4.19. <i>Confusion Matrix Percobaan Kedua</i> .....	59
Tabel 4.20. <i>Confusion Matrix Percobaan Ketiga</i> .....	60
Tabel 4.21. Hasil Percobaan .....	60
Tabel 4.22. Pengujian Sistem .....	63

## DAFTAR GAMBAR

Gambar 2.1. Proses Data Mining.....	9
Gambar 2.1. <i>Life Cycle Crisp-DM</i> .....	12
Gambar 2.2. Persamaan Teorema <i>Bayes</i> .....	14
Gambar 2.3. Teorema <i>Bayes</i> .....	14
Gambar 2.4. Teorema <i>Bayes</i> .....	15
Gambar 2.5. <i>Confusion Matrix</i> .....	17
Gambar 3.1. Alur Penelitian .....	20
Gambar 4.1. <i>Dataset Awal</i> .....	28
Gambar 4.2. Sebelum <i>Transformasi</i> .....	29
Gambar 4.3. Setelah <i>Transformasi</i> .....	29
Gambar 4.4. <i>Library Dan File Di Import</i> .....	55
Gambar 4.5. Data Dimasukan Ke <i>Data Test Dan Data Training</i> .....	55
Gambar 4.6. Hasil Prediksi Menggunakan Python .....	56
Gambar 4.7. Data Test Dan Data Training 70% Dan 30% .....	56
Gambar 4.8. Hasil Dan Akurasi Data Test Dan Data Training .....	57
Gambar 4.9. Data Test Dan Data Training 80% dan 20% .....	57
Gambar 4.10. Evaluasi Dan Hasil.....	58
Gambar 4.11. Halaman Home .....	61
Gambar 4.12. Halaman Tentang .....	61
Gambar 4.13. Halaman <i>Input Data</i> .....	62
Gambar 4.14. Halaman <i>Output Data</i> .....	62

## DAFTAR LAMBANG

No.	Gambar lambang	Nama	Keterangan
1.		<i>Start / End</i>	lambang oval yang menggambarkan dimulainya atau berhenti suatu diagram alir.
2.		<i>Flow</i>	lambang garis dengan tanda panah yang menggambarkan aliran data atau prosedur dari satu tahap ke tahap yang lain.
3.		<i>Process</i>	lambang berbentuk persegi panjang yang menggambarkan proses atau kegiatan apa yang sedang terjadi.
5.		<i>Decisions</i>	lambang berbentuk belah ketupat yang menggambarkan skenario "IF ELSE" dengan luaran <i>BOOLEAN</i> , yaitu <i>YES/NO</i> . Setiap pilihan luaran akan mengarahkan ke tahap lain.

## DAFTAR LAMPIRAN

**Lampiran 1. Biodata**

**Mahasiswa .....**

**Lampiran 2. Hasil Turnitin .....**