



**IMPLEMENTASI *GAUSSIAN NAÏVE BAYES* DENGAN OPTIMASI
PEMBOBOTAN *HILL CLIMBING* TERHADAP KLASIFIKASI
PENYAKIT KANKER PARU-PARU**

Skripsi

**Untuk memenuhi persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh
ALVIS JAUZA PRISMA
1911016310011**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2024

SKRIPSI

IMPLEMENTASI *GAUSSIAN NAÏVE BAYES* DENGAN OPTIMASI
PEMBOBOTAN *HILL CLIMBING* TERHADAP KLASIFIKASI
PENYAKIT KANKER PARU-PARU

Oleh:


ALVIS JAUZA PRISMA

1911016310011


Telah dipertahankan di depan Dosen Penguji pada tanggal 28 Juni 2024

Susunan Penguji :


Pembimbing I,


Irwan Budiman, S.T., M.Kom
NIP. 197703252008121001


Dosen Penguji I,


Dwi Kartini, S.Kom., M.Kom.
NIP. 198704212012122003

Pembimbing II,


Fatma Indriani, S.T., M.I.T., Ph.D
NIP. 198404202008122004


Dosen Penguji II,


Muhammad Itqan Mazdadi, S.Kom.,
M.Kom.
NIP. 199006122019031013



Banjarbaru, 1 Juli 2024

Ketua Program Studi Ilmu Komputer


Irwan Budiman, S.T., M.Kom
NIP. 197703252008121001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 28 Juni 2024

Yang Menyatakan,



Alvis Jauza Prisma

NIM. 1911016310011

ABSTRAK

IMPLEMENTASI *GAUSSIAN NAÏVE BAYES* DENGAN OPTIMASI PEMBOBOTAN *HILL CLIMBING* TERHADAP KLASIFIKASI PENYAKIT KANKER PARU-PARU

(Oleh : Alvis Jauza Prisma; Pembimbing: Irwan Budiman, S.T., M.Kom. dan Fatma Indriani, S.T, M.I.T., Ph.D; 2024; 58 halaman)

Penyakit kanker paru-paru merupakan salah satu penyebab utama kematian di seluruh dunia. Menurut *Globocan* pada tahun 2020 menyebutkan jumlah kanker paru-paru di Indonesia berada pada urutan ke-3 setelah kanker payudara dan kanker serviks. Dalam penelitian mengimplementasikan metode klasifikasi *Gaussian Naïve Bayes* dan dengan penambahan pembobotan dengan menggunakan metode *Hill Climbing* untuk klasifikasi penyakit kanker paru-paru. Hasil penelitian yang didapatkan dari metode *Gaussian Naïve Bayes* diperoleh nilai AUC 93,75% dan metode *Gaussian Naïve Bayes* dengan optimasi pembobotan *Hill Climbing* diperoleh nilai AUC 93,86%. Berdasarkan hasil penelitian tersebut, dapat disimpulkan metode *Gaussian Naïve Bayes* dengan optimasi pembobotan *Hill Climbing* memperoleh nilai AUC yang lebih baik daripada metode *Gaussian Naïve Bayes* saja.

Kata Kunci: Klasifikasi, pembobotan, *Gaussian Naive Bayes*, *Hill Climbing*

ABSTRACT

IMPLEMENTATION OF GAUSSIAN NAÏVE BAYES WITH HILL CLIMBING WEIGHTING OPTIMIZATION ON THE CLASSIFICATION OF LUNG CANCER DISEASE

(By : Alvis Jauza Prisma; Supervisor: Irwan Budiman, S.T., M.Kom. and Fatma Indriani, S.T, M.I.T., Ph.D; 2024; 58 pages)

Lung cancer is one of the main causes of death throughout the world. According to Globocan in 2020, the number of lung cancer in Indonesia is in 3rd place after breast cancer and cervical cancer. In this research implemented the Gaussian Naïve Bayes classification method and added weighting using the Hill Climbing method for the classification of lung cancer. The result obtained from the Gaussian Naïve Bayes method obtained an AUC value of 93,75% and the Gaussian Naïve Bayes with Hill Climbing weighting optimization obtained an AUC value of 93,86%. Based on the results, can be concluded that the Gaussian Naïve Bayes with Hill Climbing weighting optimization obtained a better AUC value than the Gaussian Naïve Bayes method alone.

Keywords: *Classification, weighting, Gaussian Naïve Bayes, Hill Climbing*

PRAKATA

Puji syukur penulis panjatkan kepada Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “IMPLEMENTASI *GAUSSIAN NAÏVE BAYES* DENGAN OPTIMASI PEMBOBOTAN *HILL CLIMBING* TERHADAP PENYAKIT KANKER PARU-PARU” untuk memenuhi syarat dalam menyelesaikan pendidikan program Srata-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut:

1. Kedua orang tua, Purwantoro dan Rustianti Amalia. Terima kasih atas do'a dan dukungan dalam menyelesaikan skripsi ini.
2. Bapak Irwan Budiman, S.T., M.Kom. selaku dosen pembimbing utama dan Ketua Program Studi Ilmu Komputer FMIPA ULM yang telah membantu dan meluangkan waktu demi kelancaran dalam menyelesaikan skripsi ini. Atas izin dari beliau jugalah skripsi ini dapat diselesaikan.
3. Fatma Indriani, S.T, M.I.T., Ph.D selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam menyelesaikan skripsi ini.
4. Seluruh Dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
5. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2019 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
6. Ucapan terima kasih kepada Farros, Falah, Amet, Chris Aldi, Fauzan Nafiz, Amir, dan teman-teman yang memberikan dukungan dalam proses mengerjakan skripsi ini.
7. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarmasin, 26 Maret 2024

A handwritten signature in black ink, appearing to read 'Alvis', written in a cursive style.

Alvis Jauza Prisma

DAFTAR ISI

PERNYATAAN	iii
ABSTRAK	iv
ABSTRACT	v
PRAKATA	vi
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian.....	2
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA	4
2.1 Kajian Terdahulu.....	4
2.2 Landasan Teori.....	9
2.2.1 Penyakit Kanker Paru-Paru	9
2.2.2 <i>Data Mining</i>	9
2.2.3 <i>Preprocessing</i>	9
2.2.4 Klasifikasi.....	10
2.2.5 <i>Gaussian Naïve Bayes Classification</i>	10
2.2.6 <i>Hill-Climbing</i>	13
2.2.7 <i>Confusion Matrix</i>	14
2.2.8 <i>Area Under Curve (AUC)</i>	15
BAB III METODE PENELITIAN	16
3.1 Alat Penelitian	16
3.2 Bahan Penelitian.....	16
3.3 Variabel Penelitian	18

3.4	Prosedur Penelitian.....	18
BAB IV HASIL DAN PEMBAHASAN.....		21
4.1	Hasil	21
4.1.1	Pengumpulan Data	21
4.1.2	<i>Preprocessing</i>	22
4.1.3	Pembagian Data.....	23
4.1.4	Pengklasifikasian.....	23
4.1.4.1	Gaussian Naïve Bayes.....	24
4.1.4.2	<i>Gaussian Naïve Bayes</i> dengan Pembobotan <i>Hill Climbing</i>	31
4.1.4.3	<i>Gaussian Naive Bayes</i> dengan Pembobotan <i>Hill Climbing</i> pada <i>Random Seed 1</i>	33
4.1.4.4	<i>Gaussian Naive Bayes</i> dengan Pembobotan <i>Hill Climbing</i> pada <i>Random Seed 2</i>	40
4.1.4.5	<i>Gaussian Naive Bayes</i> dengan Pembobotan <i>Hill Climbing</i> pada <i>Random Seed 3</i>	46
4.2	Pembahasan.....	53
BAB V PENUTUP.....		58
5.1	Kesimpulan.....	58
5.2	Saran.....	58
DAFTAR PUSTAKA		59
LAMPIRAN.....		64

DAFTAR TABEL

Tabel 1. Keaslian Penelitian.....	6
Tabel 2. Perancangan Penelitian	8
Tabel 3. <i>Confusion Matrix</i>	14
Tabel 4. Kategori klasifikasi nilai AUC.....	15
Tabel 5. Data <i>Lung Cancer</i>	17
Tabel 6. Pembagian Distribusi Orang Terkena Penyakit pada Data <i>Lung Cancer</i>	21
Tabel 7. Data <i>Lung Cancer</i>	21
Tabel 8. Data <i>Lung Cancer</i> Setelah di <i>Preprocessing</i>	22
Tabel 9. Tabel Pembagian Data	23
Tabel 10. Data training perhitungan <i>Gaussian Naïve Bayes</i>	24
Tabel 11. Nilai rata-rata tiap variable kelas 0 <i>Gaussian Naïve Bayes</i>	24
Tabel 12. Nilai rata-rata tiap variable kelas 1 <i>Gaussian Naïve Bayes</i>	25
Tabel 13. Nilai <i>variance</i> tiap variable kelas 0 <i>Gaussian Naïve Bayes</i>	25
Tabel 14. Nilai <i>variance</i> tiap variable kelas 1 <i>Gaussian Naïve Bayes</i>	26
Tabel 15. Data <i>testing</i> ke-1	26
Tabel 16. Nilai perhitungan <i>testing</i> kelas 1 <i>Gaussian Naïve Bayes</i>	26
Tabel 17. Nilai perhitungan <i>testing</i> kelas 0 <i>Gaussian Naïve Bayes</i>	27
Tabel 18. Hasil Klasifikasi <i>Gaussian Naïve Bayes</i>	27
Tabel 19 <i>Confusion Matrix</i> klasifikasi <i>Gaussian Naïve Bayes</i>	28
Tabel 20. Data Aktual dan Prediksi serta probabilitas data prediksi.	28
Tabel 21. Data Aktual dan Prediksi berdasarkan probabilitas data prediksi dari yang tertinggi	28
Tabel 22. <i>Confusion Matrix</i> pada setiap <i>threshold</i>	29
Tabel 23. <i>TPR</i> dan <i>FPR</i>	30
Tabel 24. Nilai AUC disetiap <i>Threshold</i>	31
Tabel 25. Bobot awal <i>random seed</i> 1.....	33
Tabel 26. Nilai rata-rata tiap variable kelas 0 bobot awal <i>random seed</i> 1	34
Tabel 27. Nilai rata-rata tiap variable kelas 1 bobot awal <i>random seed</i> 1	34

Tabel 28. Nilai <i>variance</i> tiap variable kelas 0 bobot awal <i>random seed</i> 1	35
Tabel 29. Nilai <i>variance</i> tiap variable kelas 1 bobot awal <i>random seed</i> 1	35
Tabel 30. Hasil Klasifikasi bobot awal <i>random seed</i> 1.....	36
Tabel 31. Nilai AUC bobot awal <i>Random Seed</i> 1.....	36
Tabel 32. Bobot kedua <i>random seed</i> 1.....	36
Tabel 33. Nilai rata-rata tiap variable kelas 0 bobot kedua <i>random seed</i> 1	37
Tabel 34. Nilai rata-rata tiap variable kelas 1 bobot kedua <i>random seed</i> 1	37
Tabel 35. Nilai <i>variance</i> tiap variable kelas 0 bobot kedua <i>random seed</i> 1	38
Tabel 36. Nilai <i>variance</i> tiap variable kelas 1 bobot kedua <i>random seed</i> 1	38
Tabel 37. Hasil Klasifikasi bobot kedua <i>random seed</i> 1.....	39
Tabel 38. Nilai AUC bobot kedua <i>Random Seed</i> 1.....	39
Tabel 39. Perbandingan nilai AUC pada bobot awal dan kedua <i>Random Seed</i> 1	39
Tabel 40. Bobot awal <i>random seed</i> 2.....	40
Tabel 41. Nilai rata-rata tiap variable kelas 0 bobot awal <i>random seed</i> 2.....	40
Tabel 42. Nilai rata-rata tiap variable kelas 1 bobot awal <i>random seed</i> 2.....	41
Tabel 43. Nilai <i>variance</i> tiap variable kelas 0 bobot awal <i>random seed</i> 2	41
Tabel 44. Nilai <i>variance</i> tiap variable kelas 1 bobot awal <i>random seed</i> 2	42
Tabel 45. Hasil Klasifikasi bobot awal <i>random seed</i> 2.....	42
Tabel 46. Nilai AUC bobot awal <i>Random Seed</i> 2.....	43
Tabel 47. Bobot kedua <i>random seed</i> 2.....	43
Tabel 48. Nilai rata-rata tiap variable kelas 0 bobot kedua <i>random seed</i> 2.....	43
Tabel 49. Nilai rata-rata tiap variable kelas 1 bobot kedua <i>random seed</i> 2	44
Tabel 50. Nilai <i>variance</i> tiap variable kelas 0 bobot kedua <i>random seed</i> 2	44
Tabel 51. Nilai <i>variance</i> tiap variable kelas 1 bobot kedua <i>random seed</i> 2	45
Tabel 52. Hasil Klasifikasi bobot kedua <i>random seed</i> 2.....	45
Tabel 53. Nilai AUC bobot kedua <i>Random Seed</i> 2.....	46
Tabel 54. Perbandingan nilai AUC pada bobot awal dan kedua <i>Random Seed</i> 2	46
Tabel 55. Perbandingan nilai AUC <i>Random Seed</i> 1 dan 2	46
Tabel 56. Bobot awal <i>random seed</i> 3.....	47

Tabel 57. Nilai rata-rata tiap variable kelas 0 bobot awal <i>random seed 3</i>	47
Tabel 58. Nilai rata-rata tiap variable kelas 1 bobot awal <i>random seed 3</i>	48
Tabel 59. Nilai <i>variance</i> tiap variable kelas 0 bobot awal <i>random seed 3</i>	48
Tabel 60. Nilai <i>variance</i> tiap variable kelas 1 bobot awal <i>random seed 3</i>	49
Tabel 61. Hasil Klasifikasi bobot awal <i>random seed 3</i>	49
Tabel 62. Nilai AUC bobot awal <i>Random Seed 3</i>	49
Tabel 63. Bobot kedua <i>random seed 3</i>	50
Tabel 64. Nilai rata-rata tiap variable kelas 0 bobot kedua <i>random seed 3</i>	50
Tabel 65. Nilai rata-rata tiap variable kelas 1 bobot kedua <i>random seed 3</i>	51
Tabel 66. Nilai <i>variance</i> tiap variable kelas 0 bobot kedua <i>random seed 3</i>	51
Tabel 67. Nilai <i>variance</i> tiap variable kelas 1 bobot kedua <i>random seed 3</i>	52
Tabel 68. Hasil Klasifikasi bobot kedua <i>random seed 3</i>	52
Tabel 69. Nilai AUC bobot kedua <i>Random Seed 3</i>	52
Tabel 70. Perbandingan nilai AUC pada bobot awal dan kedua <i>Random Seed 3</i>	53
Tabel 71. Perbandingan nilai AUC <i>Random Seed 2</i> dan 3	53
Tabel 72. Hasil Evaluasi AUC <i>Gaussian Naïve Bayes</i>	54
Tabel 73. Hasil Evaluasi AUC terbaik <i>Gaussian Naïve Bayes-Hill Climbing</i> ..	54
Tabel 74. Hasil AUC pada iterasi pembobotan <i>Hill Climbing</i>	55
Tabel 75. Hasil Evaluasi AUC terbaik <i>Gaussian Naïve Bayes-Hill Climbing</i> ..	56

DAFTAR GAMBAR

Gambar 1. Alur perhitungan metode <i>Gaussian Naive Bayes</i>	11
Gambar 2. Alur <i>Hill-Climbing</i>	13
Gambar 3. Alur Penelitian.....	18
Gambar 4. Diagram iterasi <i>Hill Climbing</i> dalam <i>Random Seed</i>	32
Gambar 5. Diagram iterasi <i>Hill Climbing</i> untuk <i>Random Seed</i> terbaik.....	33
Gambar 6. Grafik perbandingan nilai AUC	57

DAFTAR LAMPIRAN

Lampiran

Lampiran 1 *Load Library*

Lampiran 2 *Preprocessing data*

Lampiran 3 *Load dan Split data*

Lampiran 4 Model *Gaussian Naïve Bayes*

Lampiran 5 Import library tambahan untuk *Hill Climbing*

Lampiran 6 Membuat fungsi evaluasi model *Gaussian Naïve Bayes* dengan pembobotan *Hill Climbing*

Lampiran 7 Membuat model *Hill Climbing*

Lampiran 8 *Gaussian Naïve Bayes-Hill Climbing*