



**PERBANDINGAN PERFORMA *DISTANCE MEASURES* PADA
ALGORITMA *NEAREST CENTROID NEIGHBOR* DAN *K-NEAREST
NEIGHBOR* DALAM PROSES KLASIFIKASI BINTANG**

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
NURUL AZIZAH
NIM. 2111017320011**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
MEI 2025**



**PERBANDINGAN PERFORMA *DISTANCE MEASURES* PADA
ALGORITMA *NEAREST CENTROID NEIGHBOR* DAN *K-NEAREST
NEIGHBOR* DALAM PROSES KLASIFIKASI BINTANG**

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
NURUL AZIZAH
NIM. 2111017320011**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
MEI 2025**


SKRIPSI

**PERBANDINGAN PERFORMA *DISTANCE MEASURES* PADA
ALGORITMA *NEAREST CENTROID NEIGHBOR* DAN *K-NEAREST
NEIGHBOR* DALAM PROSES KLASIFIKASI BINTANG**


Oleh:
Nurul Azizah
2111017320011

Telah dipertahankan pada hari Rabu, tanggal 07-05-2025 dan disetujui oleh dosen pembimbing dan dosen penguji sebagai berikut:


Pembimbing I


Dewi Sri Susanti, S.Si., M.Si
NIP. 197305161999032002

Penguji I


Fuad Muhajirin Farid, S.Pd., M.Si
NIP. 198807112019031014

Pembimbing II


Selvi Annisa, S.Si., M.Si
NIP. 199212262022032016

Penguji II


Rifqi Aulya Rahman, S.Mat., M.Si.
NIP. 199708072024061003

Banjarbaru, 9 Mei 2025

Mengetahui,
Koordinator PS Statistika
MIPA ULM



Prof. Dewi Anggraini S.Si., M.App.Sci., Ph.D
NIP. 198303282005012001

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, 29 April 2025



Nurul Azizah
NIM. 2111017320011

PRODI STATISTIKA

ABSTRAK

Perbandingan Performa *Distance Measures* pada Algoritma *Nearest Centroid Neighbor* dan *k-Nearest Neighbor* dalam Proses Klasifikasi Bintang (Oleh: Nurul Azizah; Pembimbing: Dewi Sri Susanti dan Selvi Annisa, 2025; 67 halaman)

Bintang merupakan salah satu benda langit yang dapat diklasifikasikan berdasarkan temperatur, perbandingan luminositas bintang dengan luminositas rata-rata matahari, perbandingan jari-jari bintang dengan jari-jari rata-rata matahari, magnitudo atau kecerahan, warna umum, serta kategori *spectral* dari bintang tersebut. Bintang dapat diklasifikasikan ke dalam enam kategori, yaitu *brown dwarf*, *red dwarf*, *white dwarf*, *main sequence*, *supergiants*, dan *hypergiants*. Klasifikasi bintang atau *stellar classification* ini penting bagi astronom sebagai acuan jika ditemukan bintang jenis baru, serta untuk memahami komposisi, temperatur, dan tahap evolusi bintang. Proses klasifikasi dapat dilakukan menggunakan algoritma *Nearest Centroid Neighbor* (NCN) dan *k-Nearest Neighbor* (k-NN) dengan pendekatan jarak terdekat, seperti jarak *Euclidean*, *Manhattan*, *Minkowski*, *Chebyshev*, *Cosine*, *Jaccard*, dan *Hamming*. Berdasarkan hal tersebut dilakukan perbandingan performa algoritma NCN dan k-NN pada tujuh *distance measures*. Hasil penelitian menunjukkan bahwa klasifikasi bintang menggunakan algoritma NCN dan k-NN dengan jarak *Euclidean*, *Manhattan*, dan *Minkowski* mencapai performa sempurna, yaitu 100%. Namun, berbeda dengan jarak *Chebyshev*, yang menunjukkan performa sempurna pada algoritma k-NN, tetapi hanya mendekati sempurna pada algoritma NCN sebesar 92%. Dengan demikian, algoritma k-NN memberikan performa yang lebih unggul dibandingkan algoritma NCN dalam proses klasifikasi bintang.

Kata Kunci: Klasifikasi Bintang, *Distance Measures*, *Nearest Centroid Neighbor*, *k-Nearest Neighbor*

ABSTRACT

Comparison of Distance Measures Performance on Nearest Centroid Neighbor and k-Nearest Neighbor Algorithms in the Stellar Classification Process (By: Nurul Azizah; Supervisors: Dewi Sri Susanti dan Selvi Annisa, 2025; 67 page)

Stars are celestial bodies that can be classified based on several characteristics, including temperature, luminosity, radius, magnitude, stellar color, and spectral class. Stars are generally grouped into six categories: brown dwarfs, red dwarfs, white dwarfs, main sequence stars, supergiants, and hypergiants. Stellar classification is important to astronomers because it can help identify new types of stars and improve our understanding of their composition, temperature, and evolutionary stages. This classification process can be done using the Nearest Centroid Neighbor (NCN) and k-Nearest Neighbor (k-NN) algorithms, by applying various distance measures such as Euclidean, Manhattan, Minkowski, Chebyshev, Cosine, Jaccard, and Hamming. This study aims to compare the performance between NCN and k-NN using these seven distance measures. The results show that Euclidean, Manhattan, and Minkowski distances produce a perfect performance of 100% in both algorithms. Chebyshev distance yielded perfect performance in k-NN but slightly lower in NCN with a performance of 92%. Thus, the k-NN algorithm provides superior performance compared to the NCN algorithm in the stellar classification process.

Keywords: Stellar Classification, Distance Measures, Nearest Centroid Neighbor, k-Nearest Neighbor

PRAKATA

Puji dan syukur dipanjatkan kepada Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Performa *Distance Measures* pada Algoritma *Nearest Centroid Neighbor* dan *k-Nearest Neighbor* dalam Proses Klasifikasi Bintang”. Penyusunan skripsi ini bertujuan untuk memenuhi salah satu syarat dalam rangka menyelesaikan program sarjana di Program Studi S-1 Statistika FMIPA ULM.

Dalam penulisan skripsi, penulis banyak memperoleh bantuan, baik pengajaran, bimbingan, maupun arahan dari berbagai pihak, secara langsung maupun tidak langsung. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua dan keluarga yang selalu memberikan doa, dukungan, dan motivasi.
2. Ibu Dewi Sri Susanti, S.Si., M.Si. dan Ibu Selvi Annisa, S.Si., M.Si. selaku dosen pembimbing yang telah memberikan bimbingan, arahan, dan dukungan selama proses penyusunan skripsi.
3. Bapak Fuad Muhajirin Farid, S.Pd., M.Si. dan Bapak Rifqi Aulya Rahman, S.Mat., M.Si. selaku dosen penguji yang telah memberikan saran dan bantuan dalam perbaikan penulisan skripsi.
4. Ibu Hj. Maisarah, S.Pd.I., M.Pd. selaku dosen pembimbing akademik yang telah memberikan bimbingan dan motivasi hingga proses penyusunan skripsi.
5. Seluruh dosen dan staf Program Studi S-1 Statistika FMIPA ULM yang telah membantu dalam proses penyelesaian skripsi.
6. Teman-teman penulis yang selalu mendukung, memotivasi, dan menemani hingga proses penyusunan skripsi.
7. Seluruh pihak yang terlibat, baik secara langsung maupun tidak langsung

Penulisan skripsi ini masih terdapat kekurangan, sehingga sangat diharapkan kritik dan saran yang membangun agar dapat bermanfaat bagi semua pihak.

Banjarbaru, 29 April 2025

Nurul Azizah

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
PERNYATAAN	iii
ABSTRAK	iv
ABSTRACT	v
PRAKATA	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian	2
BAB II TINJAUAN PUSTAKA	3
2.1 Kajian Penelitian Terdahulu	3
2.2 Kajian Teori	4
2.2.1 <i>Preprocessing Data</i>	4
2.2.2 Statistika Deskriptif	6
2.2.3 <i>k-Nearest Neighbor</i>	6
2.2.4 <i>Nearest Centroid Neighbor</i>	7
2.2.5 <i>Distance Measures</i>	8
2.2.6 Pengukuran Performa Kinerja Algoritma	9
2.2.7 Bintang	11
BAB III METODE PENELITIAN	14
3.1 Sumber Data	14
3.2 Variabel Penelitian	14
3.3 Prosedur Penelitian	15
BAB IV HASIL DAN PEMBAHASAN	18
4.1 <i>Data Cleaning</i>	18
4.2 Statistika Deskriptif	19
4.3 <i>Splitting Data</i>	21
4.4 Normalisasi Variabel Numerik	21
4.5 <i>One-hot Encoding</i> Variabel Kategorik	22
4.6 Evaluasi Performa Algoritma <i>Nearest Centroid Neighbor</i>	22
4.6.1 Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Euclidean</i>	23
4.6.2 Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Manhattan</i>	26
4.6.3 Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Minkowski</i>	29
4.6.4 Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Chebyshev</i>	32

4.6.5 Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Cosine</i>	35
4.6.6 Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Jaccard</i>	40
4.6.7 Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Hamming</i>	43
4.7 Evaluasi Performa Algoritma <i>k-Nearest Neighbor</i>	47
4.7.1 Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Euclidean</i>	47
4.7.2 Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Manhattan</i>	50
4.7.3 Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Minkowski</i>	51
4.7.4 Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Chebyshev</i>	53
4.7.5 Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Cosine</i>	55
4.7.6 Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Jaccard</i>	57
4.7.7 Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Hamming</i>	59
4.8 Perbandingan Performa Algoritma <i>Nearest Centroid Neighbor</i> dan <i>k-Nearest Neighbor</i>	61
BAB V PENUTUP	64
5.1 Kesimpulan	64
5.2 Saran	64
DAFTAR PUSTAKA	66
LAMPIRAN	68
RIWAYAT HIDUP	90

DAFTAR GAMBAR

Gambar 2.1 Diagram <i>Hertzprung-Russell</i>	12
Gambar 3.1 Prosedur Penelitian	17
Gambar 4.1 Sebaran <i>Missing Value</i> pada Variabel Penelitian	18
Gambar 4.2 Persentase Kategori pada Variabel <i>Star Type</i>	20
Gambar 4.3 Performa Algoritma NCN	62
Gambar 4.4 Performa Algoritma k-NN	63

PRODI STATISTIKA

DAFTAR TABEL

Tabel 2.1	Dataset <i>One-hot Encoding</i>	5
Tabel 2.2	Contoh <i>One-hot Encoding</i>	5
Tabel 2.3	<i>Confusion Matrix</i> untuk Klasifikasi Biner.....	9
Tabel 2.4	<i>Confusion Matrix</i> untuk Klasifikasi <i>Multiclass</i>	10
Tabel 3.1	Definisi Operasional Variabel Penelitian	14
Tabel 4.1	Analisis Deskriptif Variabel Bertipe Numerik.....	19
Tabel 4.2	Analisis Deskriptif Variabel Bertipe Kategorik.....	19
Tabel 4.3	<i>Splitting Data</i>	21
Tabel 4.4	Nilai Rata-rata (μ) dan Standar Deviasi (σ)	21
Tabel 4.5	Centroid Perhitungan Jarak <i>Euclidean</i> pada Algoritma NCN	23
Tabel 4.6	Data Testing Pertama	23
Tabel 4.7	Perbandingan Akurasi Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Euclidean</i>	25
Tabel 4.8	<i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Euclidean</i>	25
Tabel 4.9	Rekapitulasi TP, TN, FP, dan FN pada Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Euclidean</i>	25
Tabel 4.10	Performa Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Euclidean</i>	26
Tabel 4.11	<i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Manhattan</i>	28
Tabel 4.12	Rekapitulasi TP, TN, FP, dan FN Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Manhattan</i>	28
Tabel 4.13	Performa Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Manhattan</i>	29
Tabel 4.14	<i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Minkowski</i>	31
Tabel 4.15	Rekapitulasi TP, TN, FP, dan FN Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Minkowski</i>	31
Tabel 4.16	Performa Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Minkowski</i>	31
Tabel 4.17	<i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Chebyshev</i>	33
Tabel 4.18	Rekapitulasi TP, TN, FP, dan FN Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Chebyshev</i>	34
Tabel 4.19	Performa Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Chebyshev</i>	34
Tabel 4.20	Centroid Perhitungan Jarak <i>Cosine</i> pada Algoritma NCN	36
Tabel 4.21	Data Testing Pertama	36
Tabel 4.22	<i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Cosine</i>	38

Tabel 4.23 Rekapitulasi TP, TN, FP, dan FN Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Cosine</i>	38
Tabel 4.24 Performa Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Cosine</i>	39
Tabel 4.25 Centroid Perhitungan Jarak <i>Jaccard</i> pada Algoritma NCN	40
Tabel 4.26 Data Testing Pertama	40
Tabel 4.27 <i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Jaccard</i>	42
Tabel 4.28 Rekapitulasi TP, TN, FP, dan FN Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Jaccard</i>	42
Tabel 4.29 Performa Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Jaccard</i>	43
Tabel 4.30 Centroid Perhitungan Jarak <i>Hamming</i> pada Algoritma NCN	44
Tabel 4.31 Data Testing Pertama	44
Tabel 4.32 <i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Hamming</i>	45
Tabel 4.33 Rekapitulasi TP, TN, FP, dan FN Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Hamming</i>	45
Tabel 4.34 Performa Klasifikasi Bintang Menggunakan Algoritma NCN dengan Jarak <i>Hamming</i>	46
Tabel 4.35 Perbandingan Akurasi Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Euclidean</i>	47
Tabel 4.36 <i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Euclidean</i>	48
Tabel 4.37 Rekapitulasi TP, TN, FP, dan FN Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Euclidean</i>	48
Tabel 4.38 Performa Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Euclidean</i>	49
Tabel 4.39 <i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Manhattan</i>	50
Tabel 4.40 Rekapitulasi TP, TN, FP, dan FN Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Manhattan</i>	50
Tabel 4.41 Performa Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Manhattan</i>	51
Tabel 4.42 <i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Minkowski</i>	52
Tabel 4.43 Rekapitulasi TP, TN, FP, dan FN Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Minkowski</i>	52
Tabel 4.44 Performa Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Minkowski</i>	53
Tabel 4.45 <i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Chebyshev</i>	54
Tabel 4.46 Rekapitulasi TP, TN, FP, dan FN Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Chebyshev</i>	54

Tabel 4.47 Performa Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Chebyshev</i>	54
Tabel 4.48 <i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Cosine</i>	55
Tabel 4.49 Rekapitulasi TP, TN, FP, dan FN Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Cosine</i>	56
Tabel 4.50 Performa Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Cosine</i>	56
Tabel 4.51 <i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Jaccard</i>	57
Tabel 4.52 Rekapitulasi TP, TN, FP, dan FN Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Jaccard</i>	58
Tabel 4.53 Performa Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Jaccard</i>	58
Tabel 4.54 <i>Confusion Matrix</i> Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Hamming</i>	59
Tabel 4.55 Rekapitulasi TP, TN, FP, dan FN Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Hamming</i>	60
Tabel 4.56 Performa Klasifikasi Bintang Menggunakan Algoritma k-NN dengan Jarak <i>Hamming</i>	60

DAFTAR LAMPIRAN

Lampiran 1. <i>Syntax Python Preprocessing Data</i>	68
Lampiran 2. <i>Syntax Python Jarak Euclidean, Manhattan, dan Chebyshev</i>	70
Lampiran 3. <i>Syntax Python Jarak Minkowski</i>	72
Lampiran 4. <i>Syntax Python Jarak Cosine</i>	74
Lampiran 5. <i>Syntax Python Jarak Jaccard dan Hamming</i>	76
Lampiran 6. <i>Data Mentah</i>	79
Lampiran 7. <i>Data Setelah Penanganan Missing Value</i>	80
Lampiran 8. <i>Posisi Missing Value</i>	81
Lampiran 9. <i>Data Konversi Variabel X Kategorik</i>	82
Lampiran 10. <i>Data Training dan Testing</i>	83
Lampiran 11. <i>Normalisasi Data</i>	85
Lampiran 12. <i>One-hot Encoding</i>	88
Lampiran 13. <i>Prediksi Data Testing pada Setiap Distance Measures</i>	89