

**Analisis Perbandingan Klasifikasi Antara KNN Dan *Naive Bayes*  
Pada Diagnosa Penyakit Kulit Menggunakan *K-Fold Cross*  
*Validation***

**SKRIPSI**



**Oleh :**

**SHOIMA FITRA KUMALA**

**NPM. 17081010022**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL  
"VETERAN" JAWA TIMUR  
2024**

**Analisis Perbandingan Klasifikasi Antara KNN dan *Naive Bayes*  
Pada Diagnosa Penyakit Kulit menggunakan *K-Fold Cross*  
*Validation***

**SKRIPSI**

Diajukan Untuk Memenuhi sebagai Persyaratan Dalam menempuh Gelar Sarjana  
Komputer Program Studi Informatika



**Oleh :**

**SHOIMA FITRA KUMALA**

**NPM. 17081010022**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR**

**2024**

# LEMBAR PENGESAHAN SKRIPSI

**Judul** : Analisis Perbandingan Klasifikasi Antara KNN dan *Naive Bayes*  
Pada Diagnosa Penyakit Kulit menggunakan *K-Fold Cross Validation*  
**Oleh** : Shoima Fitra Kumala  
**NPM** : 17081010022

Telah Diseminarkan Dalam Ujian Skripsi Pada :  
Hari Selasa, Tanggal 21 Mei 2024  
Mengetahui


**Dosen Pembimbing**

**Dosen Penguji**

1.



1.



Eka Prakarsa Mandyartha, S.T., M.Kom

Eva Yulia Puspaningrum, S.Kom., M.Kom

NIP. 19880525 201803 1 001

NIP. 19890705 202121 2 002

2.



2.



Afina Lina Nurlaili, S.Kom., M.Kom

M. Muharrom Al Haromainy, S.Kom., M.Kom

NIP. 19931213 202203 2 010

NIP. 19950601 202203 1 006

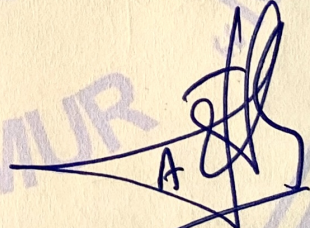
Menyetujui

**Dekan**

**Koordinator Program Studi**

**Fakultas Ilmu Komputer**

**Informatika**



Prof. Dr. Ir. Novirina Hendrasarie, MT

Fetty Tri Anggraeny, S.Kom., M.Kom

NIP. 19681126 199403 2 001

NIP. 19820211 202121 2 005

## SURAT PERNYATAAN BEBAS DARI PLAGIASI

Saya, mahasiswa Program Studi Sarjana Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Shoima Fitra Kumala

NPM : 17081010022

Menyatakan dengan sesungguhnya bahwa Skripsi/Tugas Akhir yang saya kerjakan berjudul :

“Analisis Perbandingan Klasifikasi Antara KNN dan *Naive Bayes* Pada Diagnosa Penyakit Kulit menggunakan *K-Fold Cross Validation*”

bukan merupakan plagiasi sebagian atau keseluruhan dari Skripsi/Tugas Akhir/Penelitian orang lain dari juga bukan merupakan produk dan software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/Tugas Akhir ini secara keseluruhan adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun di Institut Pendidikan lain. Bukti hasil pengecekan plagiasi dokumen ini dapat ditelusuri melalui QR Code di bawah.

Apabila di kemudian hari terbukti bahwa dokumen ini merupakan plagiasi karya orang lain, saya sanggup menerima sanksi sesuai aturan yang berlaku.

Demikian atas perhatiannya disampaikan terima kasih.

Surabaya, 21 Mei 2024

Hormat saya,



  
Shoima Fitra Kumala

NPM. 17081010022

# **ANALISIS PERBANDINGAN KLASIFIKASI ANTARA KNN DAN *NAIVE BAYES* PADA DIAGNOSA PENYAKIT KULIT MENGGUNAKAN *K-FOLD CROSS VALIDATION***

**Nama Mahasiswa : Shoima FitraKumala**

**NPM : 17081010022**

**Program Studi : Informatika**

**Dosen Pembimbing : Eka Prakarsa Mandyartha, ST., M.Kom**

**Afina Lina Nurlaili, S.Kom., M.Kom**

## **ABSTRAK**

Salah satu organ terbesar tubuh manusia ialah kulit, memiliki fungsi penting seperti perlindungan, penginderaan, dan pemeliharaan homeostasis. Meski sering dianggap kurang penting dibandingkan penyakit mematikan seperti HIV/AIDS dan tuberkulosis, masalah kulit sangat umum dan membutuhkan perhatian khusus, terutama dalam layanan kesehatan primer.

Penelitian ini membandingkan metode *Naive Bayes* dan *K-Nearest Neighbor* (KNN) dalam mendiagnosis penyakit kulit menggunakan *K-Fold cross validation*. Berdasarkan penelitian terdahulu yang menunjukkan hasil beragam dalam klasifikasi penyakit kulit, studi ini mengevaluasi akurasi kedua metode. Enam klasifikasi penyakit kulit yang diuji adalah Psoriasis, Dermatitis Seboroik, Dermatitis Kronik, Lichen Planus, Pityriasis Rosea, dan Pitiriasis Rubra Pilaris.

Hasil divalidasi data oleh 5-Fold Cross Validation sehingga menghasilkan akurasi pada *Naive Bayes* dengan nilai rata-rata akurasi 99.44% yang sebelumnya tanpa menggunakan k-fold mendapat 98.61%. Sedangkan akurasi yang di hasilkan pada algoritma *K-Nearest Neighbors* (KNN) menggunakan 5-fold cross validation adalah rata-rata akurasi 89.1% yang sebelumnya tanpa penggunaan k-fold memiliki akurasi 87.5%. Sehingga metode *Naive Bayes* mendapatkan akurasi lebih tinggi daripada *K-Nearest Neighbor*

Sehingga kesimpulan pada penelitian ini menunjukkan bahwa algoritma *Naive Bayes* memberikan klasifikasi yang sangat baik dalam mengidentifikasi penyakit kulit dibandingkan dengan KNN, dengan akurasi rata-rata lebih tinggi

setelah divalidasi data oleh *k-fold cross validation* pada dataset yang digunakan. Dengan demikian, Naive Bayes menjadi solusi yang efektif untuk permasalahan identifikasi penyakit kulit. Serta hasil penelitian menunjukkan bahwa K-Fold Cross Validation efektif dalam mengurangi overfitting dan meningkatkan validitas hasil.

**Kata Kunci** : *Naive Bayes, K-Nearest Neighbor, K-Fold Cross Validation, Data Mining, Skin Diseases, Dermatology*

## **KATA PENGANTAR**

Segala puji bagi Allah SWT. yang telah memberi penulis kemampuan untuk menyelesaikan penelitian yang berjudul "Analisis Perbandingan Klasifikasi Antara KNN dan Naive Bayes Pada Diagnosa Penyakit Kulit menggunakan K-Fold Cross Validation." Penulis senantiasa mengucapkan sholawat dan salam kepada Nabi Muhammad SAW, yang telah meneladankan ketekunan, ketelitian, dan kesabaran, sehingga penulis dapat menyelesaikan skripsi ini dengan baik.

Penulis mengucapkan terima kasih semua pihak yang telah membantu proses penulisan dari awal hingga akhir. Banyak kendala dan sumber daya yang ditemui selama proses penelitian hingga laporan ini yelesaian saat ini.

Penulis menyadari cukup banyak kesalahan dalam penelitian ini karena rendahnya tingkat pemahaman dan pengetahuan. Maka dari itu, penulis mengantisipasi dan menyambut baik kritik dan saran yang membangun dari semua pihak yang terlibat dalam pengolahan laporan ini. Saya berharap penelitian ini jelas, dan bermanfaat serta mampu menyelesaikan permasalahan banyak pihak.

Surabaya, 21 Mei 2024

Penulis,

Shoima Fitra Kumala

## UCAPAN TERIMA KASIH

Skripsi ini dilalui dengan berbagai tahapan dan rintangan yang terkait dengan pengorbanan jiwa, raga, waktu, dan harta. Banyak sekali tantangan yang dihadapi, namun dengan kegigihan dan keimanan kepada Allah subhanahu wa ta'ala, penulis mampu menyelesaikan ini dengan baik. Penulis menyatakan bahwa proses penyuntingan skripsi ini tidak bergantung pada keinginan pihak manapun yang berkepentingan.

Oleh karena itu, selama ini penulis menyampaikan rasa terima kasih yang sebesar-besarnya dan setulus-tulusnya kepada seluruh pihak yang telah memberikan dukungan berharga, baik materiil maupun emosional, baik secara terbuka maupun diam-diam, selama proses penulisan naskah ini hingga selesai, khususnya kepada:

1. Mama tercinta yaitu orang paling penting dalam hidup saya karena segala upaya dan perjuangan yang telah saya lalui hingga saat ini didukung oleh doa seorang ibu yang tidak pernah terputus setiap harinya
2. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom, sebagai koordinator program studi jurusan Informatika di Fakultas Ilmu Komputer UPN "Veteran" Jawa Timur, telah memberikan arahan dan dukungan yang sangat berarti dalam perjalanan akademik ini.
3. Bapak Eka Prakarsa Mandyartha, ST, M.Kom., sebagai Pembimbing I, telah bersedia menawarkan bimbingan dan bantuan sepanjang proses penulisan skripsi. Beliau memberikan banyak pengetahuan, serta solusi atas setiap masalah yang muncul, sehingga proses penyelesaian skripsi dapat berjalan dengan baik.
4. Ibu Afina Lina Nurlaili, S.Kom, M.Kom., sebagai Dosen Pembimbing II, telah berkenan meluangkan waktu, energi, dan pemikirannya untuk memberikan saran dan panduan selama proses penyusunan skripsi ini. Bantuan beliau sangat berharga dalam membantu penulis menyelesaikan tugas akhir ini.



5. Seluruh bapak dan ibu dosen program studi Informatika di UPN "Veteran" Jawa Timur yang telah memberikan ilmu dan pengetahuan mereka selama masa studi penulis. Berkat bimbingan dan pengajaran mereka, penulis dapat mengumpulkan pengetahuan yang diperlukan untuk melakukan penelitian ini dengan baik.
6. Seluruh teman-teman Informatika Angkatan 2017 khususnya mereka yang selalu menemani, membantu, menyemangati, mendorong, dan mendukung penulis dikala mengalami kesulitan dalam menyelesaikan penelitian.
7. Jake Enhypen, Felix Stray Kids, DK, Joshua Seventeen dan Bangtan atau musisi lainnya selaku sumber inspirasi yang mempunyai makna secara emosionalitas bagi penulis melalui karya-karyanya
8. Setiap pihak yang tidak tersebut namanya dimana penulis tidak boleh mengabaikan segala bentuk dukungan dan bantuan selama proyek penelitian ini.

# DAFTAR ISI

SKRIPSI.....	i
LEMBAR PENGESAHAN SKRIPSI.....	iii
SURAT PERNYATAAN BEBAS DARI PLAGIASI.....	iv
ABSTRAK .....	v
KATA PENGANTAR.....	vii
UCAPAN TERIMA KASIH.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR .....	xiv
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	4
1.3. Tujuan.....	4
1.4. Manfaat.....	4
1.5. Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA .....	6
2.1 Penelitian Terdahulu .....	6
2.2 Penyakit Kulit.....	9
2.3 Data Mining.....	15
2.4 Naive Bayes.....	16
2.5 K-Nearest Neighbor .....	19
2.6 Confussion Matrix.....	22

2.7	K-Fold Cross Validation .....	26
2.8	Python.....	27
BAB III METODOLOGI PENELITIAN .....		29
3.1	Tahapan Penelitian .....	29
3.2	Pengumpulan Data .....	32
3.3	Pemrosesan Data KNN.....	35
3.3.1	Load Dataset .....	35
3.3.2	Preprocessing Data .....	36
3.3.3	K-Fold Cross Validation.....	37
3.3.4	Pilih K Terbaik .....	40
3.3.5	Train - Test Split.....	40
3.3.6	Train KNN dengan k terbaik .....	41
3.3.7	Pengujian Data.....	41
3.3.8	Evaluasi Data .....	42
3.4	Pemrosesan Data Naive Bayes .....	43
3.4.1	Load Dataset .....	43
3.4.2	Preprocessing Data .....	44
3.4.3	Splitting Data.....	45
3.4.4	Train Naive Bayes .....	46
3.4.5	Pengujian Data.....	48
3.4.6	K-Fold Cross Validation.....	49
3.4.7	Evaluasi Data .....	50
BAB IV HASIL DAN PEMBAHASAN .....		53
4.1	K-Nearest Neighbor .....	53
4.1.1	Load Dataset .....	53

4.1.2	Preprocessing Data .....	55
4.1.3	K-Fold Cross Validation.....	62
4.1.4	Pilih K Terbaik .....	63
4.1.5	Splitting Data.....	63
4.1.6	Train KNN dengan k terbaik .....	64
4.1.7	Pengujian Data.....	64
4.1.8	Evaluasi Data.....	65
4.2	Naive Bayes.....	66
4.2.1	Load Dataset .....	66
4.2.2	Preprocessing Data .....	68
4.2.3	Splitting Data.....	75
4.2.4	Train Naive Bayes .....	76
4.2.5	Pengujian Data.....	76
4.2.6	K-Fold Cross Validation.....	77
4.2.7	Evaluasi Data.....	78
4.3	Perbandingan Akurasi Hasil Klasifikasi Penyakit Kulit .....	79
BAB V KESIMPULAN DAN SARAN.....		81
5.1	Kesimpulan .....	81
5.2	Saran.....	81
DAFTAR PUSTAKA .....		83

## DAFTAR TABEL

Tabel 2. 1 Kajian Pustaka.....	8
Tabel 2. 2 Confusion Matrix .....	23
Tabel 3. 1 Dataset Dermatology.....	33
Tabel 3. 2 Kelas jenis penyakit kulit.....	35
Tabel 3. 3 Contoh Fold ke-1 .....	38
Tabel 3. 4 Data Latih KNN fold ke-1 .....	39
Tabel 3. 5 Jumlah Data Penyakit.....	46
Tabel 3. 6 Likelihood atribut erythema class 1 .....	47
Tabel 3. 7 Probabilitas Likelihood erythema .....	47
Tabel 3. 8 Likelihood Atribut Scaling .....	48
Tabel 3. 9 Probabilitas Likelihood Scaling .....	48
Tabel 3. 10 Contoh kelas 1 datates-52 .....	49
Tabel 3. 11 Fold ke-n .....	50
Tabel 3. 12 Contoh Data Latih Naive Bayes 286 Data .....	50
Tabel 4. 1 Hasil 5-Fold Cross Validation.....	62
Tabel 4. 2 Hasil Prediksi KNN .....	65
Tabel 4. 3 Hasil Prediksi Naive Bayes.....	77
Tabel 4. 4 Hasil akurasi tiap fold .....	78
Tabel 4. 5 Hasil Perbandingan 2 Metode .....	79

## DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi K-Fold Cross Validation .....	26
Gambar 3. 1 Diagram Alur Tahapan Penelitian.....	29
Gambar 3. 2 Diagram Alur KNN.....	35
Gambar 3. 3 Diagram Alur Naive Bayes .....	43
Gambar 4. 1 Memasang/menyambungkan google drive.....	53
Gambar 4. 2 Import Library .....	53
Gambar 4. 3 Membaca Data .....	54
Gambar 4. 4 Lanjutan Membaca Data .....	54
Gambar 4. 5 Periksa nilai yang hilang pada Atribut .....	55
Gambar 4. 6 Informasi kolom NaN.....	56
Gambar 4. 7 Menampilkan data “?” .....	56
Gambar 4. 8 Menampilkan data terhapus .....	56
Gambar 4. 9 Menampilkan banyak data disetiap kelas.....	57
Gambar 4. 10 Diagram Class Distribution.....	57
Gambar 4. 11 Periksa Nilai Null .....	58
Gambar 4. 12 Info Data Frame .....	59
Gambar 4. 13 Ubah Tipe Data .....	59
Gambar 4. 14 Ubah tipe data .....	60
Gambar 4. 15 Membagi dataset menjadi fitur dan label .....	61
Gambar 4. 16 Menampilkan data lima baris pertama X .....	61
Gambar 4. 17 Lima baris pertama y.....	61
Gambar 4. 18 Tampilan 5-Fold Cross validation pada knn .....	62
Gambar 4. 19 Grafik tingkat akurasi k pada KNN.....	63
Gambar 4. 20 Pemisahan Dataset.....	63
Gambar 4. 21 Hasil pemisahan data train sebanyak 286 Data.....	64
Gambar 4. 22 Hasil pemisahan data train sebanyak 72 Data.....	64
Gambar 4. 23 Melatih model KNN.....	64
Gambar 4. 24 Hasil prediksi penyakit.....	64
Gambar 4. 25 Hasil Metrik Evaluasi KNN .....	65
Gambar 4. 26 Memasang/menyambungkan google drive.....	66

Gambar 4. 27 Import Library .....	66
Gambar 4. 28 Membaca Data.....	67
Gambar 4. 29 Lanjutan Membaca Data .....	67
Gambar 4. 30 Periksa nilai yang hilang pada Atribut .....	68
Gambar 4. 31 Informasi kolom NaN.....	69
Gambar 4. 32 Menampilkan data “?” .....	69
Gambar 4. 33 Menampilkan data terhapus .....	69
Gambar 4. 34 Menampilkan banyak data disetiap kelas .....	70
Gambar 4. 35 Diagram Class Distribution.....	70
Gambar 4. 36 Periksa Nilai Null .....	71
Gambar 4. 37 Info Data Frame .....	72
Gambar 4. 38 Ubah Tipe Data .....	72
Gambar 4. 39 Ubah tipe data .....	73
Gambar 4. 40 Diskritisasi Data pada atribut age.....	73
Gambar 4. 41 Membagi dataset menjadi fitur dan label .....	74
Gambar 4. 42 Menampilkan data lima baris pertama X .....	74
Gambar 4. 43 Lima baris pertama y.....	75
Gambar 4. 44 Pemisahan Dataset.....	75
Gambar 4. 45 Hasil pemisahan data train sebanyak 286 Data.....	75
Gambar 4. 46 Hasil pemisahan data train sebanyak 72 Data.....	75
Gambar 4. 47 Hasil probabilitas Prior .....	76
Gambar 4. 48 Hasil Probabilitas Likelihood.....	76
Gambar 4. 49 Hasil Prediksi Naive Bayes .....	76
Gambar 4. 50 Evaluasi Metrik Naive Bayes .....	77
Gambar 4. 51 5-fold validation .....	78
Gambar 4. 52 Grafik 5-Fold Cross Validation pada Naive Bayes.....	78
Gambar 4. 53 Confussion Matrix dan Hasil metrik evaluasi naive bayes .....	79