

**KLASIFIKASI PENYAKIT POLYCYSTIC OVARIAN SYNDROME (PCOS)
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR**



**Natasha Leslie
00000041894**

UMN

**UNIVERSITAS
MULTIMEDIA
NUSANTARA**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2023

**KLASIFIKASI PENYAKIT POLYCYSTIC OVARIAN SYNDROME (PCOS)
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR**



**Natasha Leslie
00000041894**

UMMN

**UNIVERSITAS
MULTIMEDIA
NUSANTARA**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2023

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Natasha Leslie
Nomor Induk Mahasiswa : 00000041894
Program Studi : Informatika

Skripsi dengan judul:

Klasifikasi Penyakit Polycystic Ovarian Syndrome (PCOS) menggunakan Algoritma K-Nearest Neighbor

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 16 Juni 2023



(Natasha Leslie)

HALAMAN PENGESAHAN

Skripsi dengan judul

KLASIFIKASI PENYAKIT POLYCYSTIC OVARIAN SYNDROME (PCOS) MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR

oleh

Nama : Natasha Leslie
NIM : 00000041894
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Rabu, 21 Juni 2023

Pukul 08.00 s/s 10.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang

Penguji

(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)

NIDN: 0311106903

(Adityawan, S.Komp., M.Si.)

NIDK: 8994550022

Pembimbing

(Angga Aditya Permana, S.Kom., M.Kom.)

NIDN: 0407128901

Ketua Program Studi Informatika,

(Marlinda Vasty Overbeek, S.Kom., M.Kom.)

NIDN: 0818038501

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

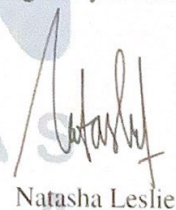
Nama : Natasha Leslie
NIM : 00000041894
Program Studi : Informatika
Fakultas : Teknik dan Informatika
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada **Universitas Multimedia Nusantara** hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

**KLASIFIKASI PENYAKIT POLYCYSTIC OVARIAN SYNDROME (PCOS)
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media / format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

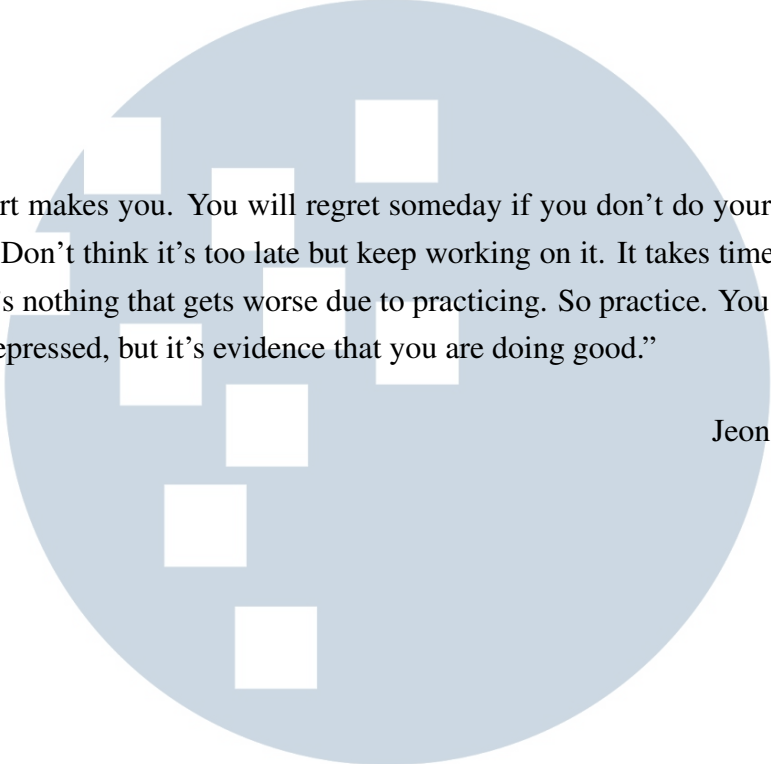
Tangerang, 16 Juni 2023
Yang menyatakan



Natasha Leslie

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Halaman Persembahan / Motto



”Effort makes you. You will regret someday if you don’t do your best now. Don’t think it’s too late but keep working on it. It takes time, but there’s nothing that gets worse due to practicing. So practice. You may get depressed, but it’s evidence that you are doing good.”

Jeon Jungkook

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

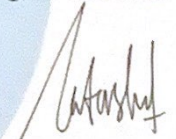
Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Klasifikasi Penyakit Polycystic Ovarian Syndrome (PCOS) menggunakan Algoritma K-Nearest Neighbor dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Angga Aditya Permana, S.Kom., M.Kom., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Orang Tua, keluarga, dan sahabat saya yang telah memberikan bantuan dukungan material dan moral, sehingga dapat menyelesaikan skripsi.
6. Rekan-rekan saya yang hingga saat ini masih banyak berkomunikasi dan banyak membantu saya yaitu Andrea, Violy, Nata, Madelinne, Jessica, Gape, Sul, Martin, Iron, Bintang, Nanda, Maureen, My HEM My Adventure Gengs, Armanda, Farrel, Elfajar, Irvan, Denisha, Hanna, Irene, Ibnu, Rafif, Ka Gaby, Susan, Zefanya, Upi gede, Bagas, Tasya, S, G, Ka Anna, dan Ka Ika yang sudah senantiasa menemani selama masa perkuliahan.
7. Kim Namjoon, Kim Seokjin, Min Yoongi, Jung Hoseok, Park Jimin, Kim Taehyung, dan Jeon Jungkook yang telah menghibur saya dari awal perjalanan kuliah hingga saat ini saya menyelesaikan skripsi saya khususnya Jeon Jungkook yang sudah menjadi tempat saya bersandar selama ini.

8. Pemilik NIM 00000037741 yang sudah membantu dan mendengar keluhan kesah saya selama menyelesaikan skripsi ini.
9. *Last but not least, I wanna thank me. I wanna thank me for believing in me. I wanna thank me for all doing this hard work. I wanna thank me for having no days off. I wanna thank me for never quitting. I wanna thank me for just being me at all times.*

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 16 Juni 2023



Natasha Leslie

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

KLASIFIKASI PENYAKIT POLYCYSTIC OVARIAN SYNDROME (PCOS) MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR

Natasha Leslie

ABSTRAK

Gangguan yang biasa dialami oleh wanita berumur subur salah satunya adalah gangguan siklus menstruasi. Penyakit Polycystic Ovarian Syndrom atau PCOS merupakan salah satu penyakit yang memiliki dampak yang signifikan pada kesehatan reproduksi wanita. PCOS masih menjadi penyakit gangguan hormon pada Wanita yang seringkali tidak disadari. Dampak dari PCOS terhadap kesehatan wanita, membuat melakukan penelitian ini akan berkontribusi untuk bisa memajukan pengetahuan bagian medis, lebih mudah menilai hasil pasien serta meningkatkan pemahaman akan penyakit PCOS. Penelitian ini mencoba mengembangkan sistem sederhana menggunakan *machine learning* untuk mengetahui status pasien berkategori terkena penyakit PCOS atau tidak dengan memasukkan beberapa informasi dari gejala yang dialami oleh pasien. Sistem dikembangkan menggunakan metode klasifikasi K-Nearest Neighbor (KNN) dan membandingkan nilai akurasi yang didapatkan dari setiap uji coba yang dilakukan untuk mendapatkan performa terbaik. Dalam percobaan ini dilakukan pembagian data dengan rasio data latih dan data uji yang berbeda-beda. Didapatkan akurasi sebesar 100% dari pembagian rasio data latih sebesar 75% dan data latih sebesar 70% dengan nilai $K=3$ dan $K=11$, dan data latih sebesar dengan nilai $K=13$ menggunakan uji validasi silang 10-fold.

Kata kunci: *K-Fold Cross Validation, K-Nearest Neighbor, Machine Learning, PCOS*



Classification of Polycystic Ovarian Syndrome (PCOS) using the K-Nearest Neighbor Algorithm

Natasha Leslie

ABSTRACT

Disorders that are commonly experienced by women at dawn, one of which is menstrual cycle disorders. Polycystic Ovarian Syndrome or PCOS is a disease that has a significant impact on women's reproductive health. PCOS is still a hormonal disorder in women that is often not realized. The impact of PCOS on women's health, conducting this research will contribute to advancing medical knowledge, more easily assessing patient outcomes and increasing understanding of PCOS disease. This study attempts to develop a simple system using *machine learning* to find out whether a patient is categorized as having PCOS or not by entering some information from the symptoms experienced by the patient. The system was developed using the K-Nearest Neighbor (KNN) classification method and compared the accuracy values obtained from each trial conducted to get the best performance. In this experiment, the distribution of data was carried out with different ratios of training data and test data. An accuracy of 100% is obtained from the distribution of the training data ratio of 75% and training data of 70% with values $K=3$ and $K=11$, and the training data is equal to the value $K=13$ using the *10-fold cross validation test*.

Keywords: *K-Fold Cross Validation, K-Nearest Neighbor, Machine Learning, PCOS*



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR KODE	xiv
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Permasalahan	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Spesifikasi Sistem	4
1.7 Sistematika Penulisan	5
BAB 2 LANDASAN TEORI	6
2.0.1 Polycystic Ovarian Syndrome (PCOS)	6
2.0.2 K-Nearest Neighbor Classifier	8
2.0.3 K-Fold Cross Validation	9
2.0.4 Confussion Matrix	10
BAB 3 METODOLOGI PENELITIAN	12
3.1 Metodologi Penelitian	12
3.1.1 Studi Literatur	12
3.1.2 Pengumpulan Data	12
3.1.3 Pengolahan Data	12
3.1.4 Pengujian dan Evaluasi	13
3.1.5 Pembuatan Laporan	13
3.2 Prosedur Penelitian	13
3.2.1 Gambaran Umum Perancangan	13
3.2.2 Raw Data	15
3.2.3 Pre-processing	15
3.2.4 Train-Test Split Data	16
3.2.5 Klasifikasi	16
3.2.6 Fitting Model	16
3.2.7 Evaluation	17
BAB 4 HASIL DAN DISKUSI	18
4.1 Implementasi Sistem	18
4.1.1 Import Dataset dan Pre-processing	18
4.1.2 Train Test Split Data	23
4.1.3 Apply KNN	24
4.1.4 Evaluasi	29

4.2	Uji Coba	35
4.2.1	Pengujian Parameter K dengan Pembagian Data Training dan Testing	36
4.2.2	Grafik Hasil Klasifikasi	41
4.2.3	Pengujian dengan Input Prediksi	44
BAB 5	SIMPULAN DAN SARAN	45
5.1	Simpulan	45
5.2	Saran	46
	DAFTAR PUSTAKA	47



DAFTAR GAMBAR

Gambar 2.1	Skema <i>10-Fold cross validation</i>	10
Gambar 3.1	Flowchart metodologi penelitian	12
Gambar 3.2	Flowchart gambaran umum perancangan	14
Gambar 3.3	Flowchart pre-processing	15
Gambar 3.4	Flowchart proses pemodelan <i>K-Nearest Neighbor</i>	17
Gambar 4.1	Hasil <i>output</i> pencarian data <i>null</i>	21
Gambar 4.2	<i>Data balance</i> atau proporsi kelas	22
Gambar 4.3	Hasil akurasi setiap <i>fold</i> rasio 90:10	25
Gambar 4.4	Hasil akurasi setiap <i>fold</i> rasio 80:20	25
Gambar 4.5	Hasil akurasi setiap <i>fold</i> rasio 75:25	26
Gambar 4.6	Hasil akurasi setiap <i>fold</i> rasio 70:30	26
Gambar 4.7	Hasil akurasi setiap <i>fold</i> rasio 60:40	27
Gambar 4.8	Hasil akurasi setiap <i>fold</i> rasio 50:50	27
Gambar 4.9	Hasil <i>confusion matrix</i> ratio 90:10	30
Gambar 4.10	Hasil <i>classification report</i> ratio 90:10	30
Gambar 4.11	Hasil <i>confusion matrix</i> ratio 80:20	31
Gambar 4.12	Hasil <i>classification report</i> ratio 80:20	31
Gambar 4.13	Hasil <i>confusion matrix</i> ratio 75:25	32
Gambar 4.14	Hasil <i>classification report</i> ratio 75:25	32
Gambar 4.15	Hasil <i>confusion matrix</i> ratio 70:30	33
Gambar 4.16	Hasil <i>classification report</i> ratio 70:30	33
Gambar 4.17	Hasil <i>confusion matrix</i> rasio 60:40	34
Gambar 4.18	Hasil <i>classification report</i> rasio 60:40	34
Gambar 4.19	Hasil <i>confusion matrix</i> rasio 50:50	35
Gambar 4.20	Hasil <i>classification report</i> rasio 50:50	35
Gambar 4.21	<i>Classification report</i> $K=3$ ratio 75:25	39
Gambar 4.22	<i>Classification report</i> $K=11$ ratio 75:25	40
Gambar 4.23	<i>Classification report</i> $K=3$ ratio 70:30	40
Gambar 4.24	<i>Classification report</i> $K=11$ ratio 70:30	40
Gambar 4.25	<i>Classification report</i> $K=13$ ratio 60:40	40
Gambar 4.26	Hasil klasifikasi PCOS	41
Gambar 4.27	Hasil persentase perbandingan antar pembagian data	42
Gambar 4.28	Hasil grafik <i>classification report</i> 90:10	43
Gambar 4.29	Hasil grafik <i>classification report</i> 80:20	43
Gambar 4.30	Hasil akhir uji input data array 1	44
Gambar 4.31	Hasil akhir uji input data array 2	44

U N I V E R S I T A S
 M U L T I M E D I A
 N U S A N T A R A

DAFTAR TABEL

Tabel 2.1	Confusion matrix	10
Tabel 4.1	Head raw dataset PCOS	19
Tabel 4.2	Head dataset PCOS	21
Tabel 4.3	Hasil uji coba parameter K	38
Tabel 4.4	Hasil akhir neighbors	39



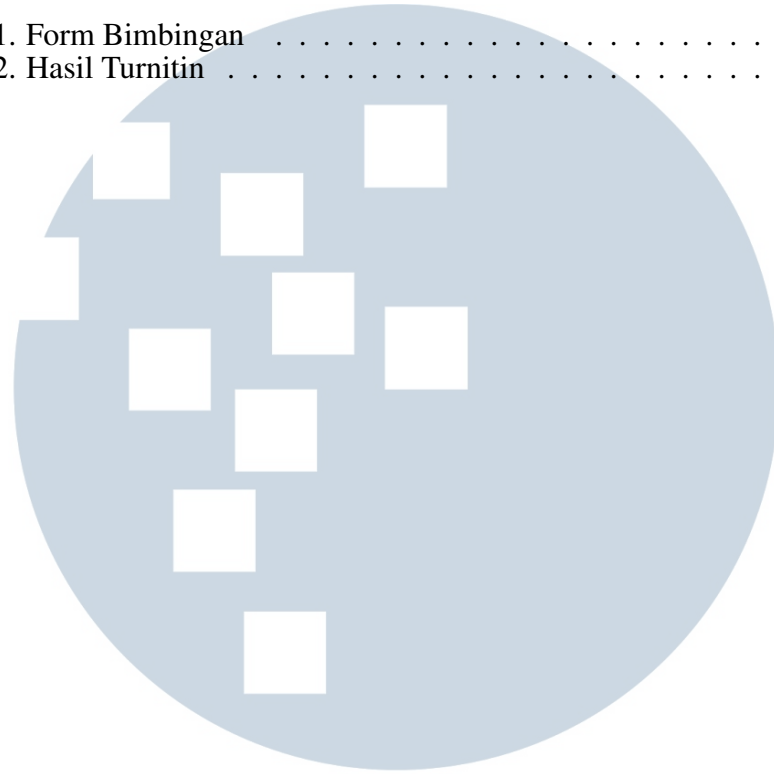
DAFTAR KODE

4.1	Potongan import libraries	18
4.2	Potongan import dan read dataset	18
4.3	Potongan kode <i>drop</i> dan <i>rename</i> kolom <i>dataset</i>	19
4.4	Potongan kode melakukan <i>encoding catagorical dataset</i> kolom periode menstruasi	19
4.5	Potongan kode melakukan <i>encoding catagorical dataset</i> kolom status	20
4.6	Potongan kode mencari data yang bersifat null	20
4.7	Potongan kode normalisasi data	21
4.8	Potongan kode pembagian data X dan Y untuk <i>training</i> dan <i>testing</i>	21
4.9	Potongan kode <i>split data</i> 90 : 10	23
4.10	Potongan kode <i>split data</i> 80 : 20	23
4.11	Potongan kode <i>split data</i> 75 : 25	23
4.12	Potongan kode <i>split data</i> 70 : 30	23
4.13	Potongan kode <i>split data</i> 60 : 40	23
4.14	Potongan kode <i>split data</i> 50 : 50	23
4.15	Potongan kode membuat objek KNN	24
4.16	Potongan kode menentukan <i>K-fold cross validation</i>	24
4.17	Potongan kode menentukan nilai K terbaik rasio 90:10	27
4.18	Potongan kode melatih KNN dengan nilai K terbaik pada data 10:80	28
4.19	Potongan kode menguji model	28
4.20	Potongan kode <i>confusion matrix</i>	29
4.21	Potongan kode uji coba parameter K	36



DAFTAR LAMPIRAN

Lampiran 1. Form Bimbingan	50
Lampiran 2. Hasil Turnitin	52



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA