

SKRIPSI
KOMPARASI ALGORITMA C4.5 DENGAN ALGORITMA
K-NEAREST NEIGHBOR (KNN) UNTUK KLASIFIKASI
STATUS GIZI BALITA



SARAS ESHANTY

NIM : 225611025

PROGRAM STUDI SISTEM INFORMASI
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA
2024

SKRIPSI
KOMPARASI ALGORITMA C4.5 DENGAN ALGORITMA
K-NEAREST NEIGHBOR (KNN) UNTUK KLASIFIKASI
STATUS GIZI BALITA

Diajukan sebagai salah satu syarat untuk menyelesaikan studi



Program Sarjana
Program Studi Sistem Informasi
Fakultas Teknologi Informasi
Universitas Teknologi Digital Indonesia
Yogyakarta

Disusun Oleh
SARAS ESHANTY
NIM : 225611025

PROGRAM STUDI SISTEM INFORMASI
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA
2024

HALAMAN PERSETUJUAN
UJIAN SKRIPSI

Judul : *Komparasi Algoritma C4.5 dengan Algoritma K-Nearest Neighbor (KNN) Untuk Klasifikasi Status Gizi Balita*

Nama : Saras Eshanty

NIM : 225611025

Program Studi : Sistem Informasi

Program : Sarjana

Semester : Gasal

Tahun Akademik : 2023/2024

Telah diperiksa dan disetujui untuk diujikan di hadapan Dewan Penguji Skripsi

Yogyakarta, 7 Pebruari 2024

Dosen Pembimbing,



Sumiyatun, S.Kom., M.Cs.

NIP/NPP : 091158

HALAMAN PENGESAHAN

SKRIPSI

KOMPARASI ALGORITMA C4.5 DENGAN ALGORITMA K-NEAREST
NEIGHBOR (KNN) UNTUK KLASIFIKASI STATUS GIZI BALITA

Telah dipertahankan di depan Dewan Penguji Skripsi dan dinyatakan
diterima untuk memenuhi sebagai persyaratan guna memperoleh Gelar

Sarjana Komputer
Program Studi Sistem Informasi
Universitas Teknologi Digital Indonesia
Yogyakarta

Yogyakarta, 7 Pebruari 2024

Dewan Penguji

NIP/NPP

Tanda

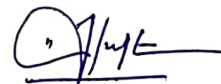
1. Danny Kriestanto, S.Kom., M.Eng. (Ketua)

0503068002

Tangan


2. Sumiyatun, S.Kom., M. Cs. (Sekretaris)

091158



Mengetahui
Ketua Program Studi Sistem Informasi



Deborah Kurniawati, S.Kom., M.Cs.

19 FEB 2024

NPP : 051149

PERNYATAAN KEASLIAN SKRIPSI

Dengan ini saya menyatakan bahwa naskah skripsi ini belum pernah diajukan untuk memperoleh gelar Sarjana Komputer di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara sah diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 7 Pebruari 2024



Saras Eshanty

NIM .225611025

DAFTAR ISI

HALAMAN JUDUL	I
HALAMAN PERSETUJUAN	III
HALAMAN PENGESAHAN	IV
PERNYATAAN KEASLIAN SKRIPSI.....	V
DAFTAR ISI.....	VI
DAFTAR GAMBAR.....	IX
DAFTAR TABEL	IX
DAFTAR LAMPIRAN	XII
KATA PENGANTAR.....	XIII
ABSTRAKSI.....	XV
BAB 1 PENDAHULUAN.....	1
1.1. LATAR BELAKANG MASALAH.....	1
1.2. RUMUSAN MASALAH.....	3
1.3. RUANG LINGKUP	3
1.4. TUJUAN PENELITIAN.....	3
1.5. MANFAAT PENELITIAN	4
1.6. SISTEMATIKA PENULISAN	4
BAB 2 TINJAUAN PUSTAKA DAN DASAR TEORI	6
2.1. TINJAUAN PUSTAKA	6
2.2. DASAR TEORI.....	8
2.2.1. <i>Definisi Data Mining</i>	<i>8</i>
2.2.2. <i>Proses Data Mining.....</i>	<i>9</i>
2.2.3. <i>Teknik Data Mining.....</i>	<i>11</i>
2.2.4. <i>Cross Industry Standard Process for Data Mining (CRISP-DM).....</i>	<i>13</i>

2.2.5. <i>Machine Learning</i>	14
2.2.6. <i>Decision Tree</i>	16
2.2.7. <i>Algoritma C4.5</i>	16
BAB 3 METODOLOGI PENELITIAN	41
3.1. TUJUAN PENELITIAN	41
3.2. WAKTU DAN TEMPAT PENELITIAN	41
3.3. ALAT DAN BAHAN PENELITIAN	41
3.3.1. <i>Alat Penelitian</i>	41
3.3.2. <i>Bahan Penelitian</i>	41
3.4. OBJEK PENELITIAN	42
3.5. METODE PENGUMPULAN DATA	42
3.6. METODE PENELITIAN	43
3.6.1 <i>Pemahaman Bisnis (Business Understanding)</i>	43
3.6.2 <i>Pemahaman Data (Data Understanding)</i>	43
3.6.3 <i>Pengolahan Data (Data Preparation)</i>	44
3.6.4 <i>Pemodelan (Modelling)</i>	45
3.6.5 <i>Evaluasi (Evaluation)</i>	45
3.6.6 <i>Deployment</i>	46
BAB 4 IMPLEMENTASI DAN PEMBAHASAN	47
4.1 IMPLEMENTASI DATA	47
4.1.1 <i>Pengumpulan Data</i>	47
4.1.2 <i>Cleaning Data</i>	47
4.1.3 <i>Seleksi Data</i>	48
4.1.4 <i>Penyiapan File</i>	48
4.2 CLASSIFICATION	48
4.2.1 <i>Penerapan Algoritma Decision Tree C4.5</i>	49
4.2.2 <i>Penerapan Algoritma K-Nearest Neighbor (KNN)</i>	58
BAB 5 SIMPULAN DAN SARAN	65
5.1. SIMPULAN	65

5.2 SARAN.....	65
DAFTAR PUSTAKA.....	66

DAFTAR GAMBAR

Gambar 2.1 Tahapan Proses KDD.....	9
Gambar 2.2 Prosedur Umum Data Mining	10
Gambar 2.3 Arsitektur Machine Learning	14
Gambar 2.4 Pohon Keputusan Hasil Perhitungan Node 1	22
Gambar 2.5 Pohon Keputusan Hasil Perhitungan Node 1.1	23
Gambar 2.6 Pohon Keputusan Hasil Perhitungan Node 1.1.2	24
Gambar 2.7 Algoritma K-Nearest Neighbor	26
Gambar 4.1 Pohon Keputusan Node 1.....	53
Gambar 4.2 Pohon Keputusan Node 1 dari RapidMiner 10.3.....	53
Gambar 4.3 Pohon Keputusan Node 1.1	55
Gambar 4.4 Pohon Keputusan Node 1.1 dengan Rapidminer 10.3	56
Gambar 4.5 Design Rapidminer 10.3 untuk Decission Tree.....	56
Gambar 4.6 Design Rapidminer 10.3 Algoritma C4.5 untuk melihat performa ...	57
Gambar 4.7 Confussion Matrix C4.5	57
Gambar 4.8 Design Algoritma KNN dari Rapidminer 10.3	63
Gambar 4.9 Confussion Matrix Algoritma KNN dari Rapidminer 10.3	64

DAFTAR TABEL

Tabel 2 .1 Keputusan Bermain Tennis.....	19
Tabel 2 .2 Perhitungan Node 1	21
Tabel 2 .3 Perhitungan Node 1.1	22
Tabel 2. 4 Perhitungan Node 1.1.2	23
Tabel 2 .5 Data Training	27
Tabel 2 .6 Perhitungan Jarak Menggunakan Euclidean Distance	29
Tabel 2 .7 Jarak Terdekat Sebanyak K=5	29
Tabel 2 .8 Indeks Massa Tubuh.....	39
Tabel 4. 1 Data Status Gizi Balita (BB/TB).....	47
Tabel 4. 2 Data Status Gizi setelah melalui tahap cleaning data dan selection data	48
Tabel 4. 3 Perhitungan Jumlah Kasus Total BB/TB.....	51
Tabel 4. 4 Perhitungan Entropy dan Gain Node 1	52
Tabel 4. 5 Perhitungan Entropy dan Gain Node 1.1	53
Tabel 4. 6 Perhitungan Entropy dan Gain Node 1.1	54
Tabel 4. 7 Perhitungan Entropy dan Gain Node 1.1	54
Tabel 4. 8 Perhitungan Entropy dan Gain Node 1.1	55
Tabel 4. 9 Data Prediksi C4.5 dari Rapidminer 10.3	56
Tabel 4. 10 Data Training	58
Tabel 4. 11 Data Testing	58
Tabel 4. 12 Kriteria Jenis Kelamin	59
Tabel 4. 13 Kriteria Usia	59
Tabel 4. 14 Kriteria BB/U	59
Tabel 4. 15 Kriteria TB/U	59
Tabel 4. 16 Konversi Data Training	60
Tabel 4. 17 Konversi Data Testing	60
Tabel 4. 18 Perhitungan Euclidean Distance.....	61
Tabel 4. 19 Hasil Perhitungan Euclidean Distance	61

Tabel 4. 20 Hasil Pengurutan dari hasil jarak terkecil ke terbesar	62
Tabel 4. 21Tabel Terdekat dari Data Uji	62
Tabel 4. 22 Penilaian Status Gizi.....	62
Tabel 4. 23 Tabel Data Prediksi KNN dari Rapidminer 10.3	64

DAFTAR LAMPIRAN

Lampiran 4. 1 Tabel Data Lengkap Status Gizi Balita (BB/TB)	68
Lampiran 4 .2 Tabel Data Lengkap Status Gizi Setelah Melalui Tahap Cleaning Data dan Selection Data	88

KATA PENGANTAR

Segala puji dan syukur Penulis panjatkan kehadirat Allah SWT, atas karunia-Nya kepada Penulis sehingga dapat menyelesaikan Tugas Akhir yang berjudul “Komparasi *Algoritma C4.5* Dengan *Algoritma K-Nearest Neighbor (KNN)* Untuk Klasifikasi Status Gizi Balita”. Shalawat serta salam Penulis panjatkan kepada junjungan kita Nabi Muhammad SAW yang kita harapkan safa’atnya di yaumul akhir.

Adapun penyusunan Tugas Akhir ini diajukan untuk memenuhi tugas akhir sebagai salah satu syarat untuk memperoleh gelar sarjana.

Dalam penyusunan Tugas Akhir ini Penulis banyak mengalami hambatan dan kesulitan akan tetapi atas bimbingan serta arahan dari para pembimbing dan dukungan semua pihak yang bersangkutan, Penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Untuk itu Penulis mengucapkan terima kasih kepada :

1. Dr. Bambang Purnomosidi DP, S.E. Akt., S. Kom., MMSI. selaku Dekan Fakultas Teknologi Informasi Universitas Teknologi Digital Indonesia (UTDI) Yogyakarta
2. Deborah Kurniawati, S.Kom., M.Cs. selaku Kaprodi Sistem Informasi Universitas Teknologi Digital Indonesia (UTDI) Yogyakarta
3. Sumiyatun, S.Kom., M.Cs. selaku Dosen Pembimbing
4. Seluruh dosen dan staf Universitas Teknologi Digital Indonesia (UTDI) Yogyakarta
5. Dinas Kesehatan Kabupaten Wonosobo yang telah mengizinkan Penulis melakukan studi kasus dan telah memberikan bimbinganya dalam penyusunan Tugas Akhir ini.
6. Seluruh keluarga yang telah memberikan dukungan, doa serta segala fasilitas dalam penyusunan Tugas Akhir ini.
7. Teman – teman Prodi SI angkatan 2022 Fakultas Teknik Informasi Program Studi Sistem Informasi Universitas Teknologi dan Digital Indonesia (UTDI) Yogyakarta.

8. Seluruh pihak yang telah membantu dalam penyusunan Tugas Akhir ini yang tidak bisa disebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini masih jauh dari sempurna. Untuk itu Penulis mengharapkan kritik dan saran yang membangun untuk perbaikan Tugas Akhir ini.

Akhir kata, Penulis berharap agar Tugas Akhir ini dapat berguna bagi semua pihak serta menjadi sesuatu yang berarti dari usaha penulis selama ini.

Wonosobo, Pebruari 2024

Penulis

ABSTRAKSI

Kesehatan anak usia 0-60 bulan harus sempurna untuk membentuk generasi emas dimasanya dan gizi merupakan asupan penting yang mempengaruhi tumbuh kembang seorang anak, kebutuhan gizi pada anak terutama balita sangat rentan mengalami perubahan berat badan. Untuk mengetahui penilaian status gizi, balita akan melakukan pemeriksaan *antropometri* yang terdiri dari tinggi badan dan berat badan secara berkala dan hasil pemeriksaan tersebut dicatat secara manual pada *KMS (Kartu Menuju Sehat)*. Metode tersebut kurang akurat serta kurang lengkap karena penilaian status gizi hanya berdasarkan berat badan menurut umur (BB/U) balita saja.

Oleh sebab itu digunakanlah *algoritma data minning* untuk *mengklasifikasi* status gizi balita, akan ditentukan jenis algoritma mana yang paling baik atau akurat untuk *mengklasifikasi* status gizi balita berdasarkan komparasi dua jenis algoritma yaitu *Algoritma C.45* dan *Algoritma K-Nearest Neighbour (KNN)* dengan menggunakan *tools Rapidminer versi 10.3* dengan mempertimbangkan 4 faktor yaitu: 1) Jenis Kelamin, 2) Usia, 3) Pengukuran Berat Badan sesuai Umur (BB/U), 4) Pengukuran Tinggi Badan (TB/U) kemudian menghasilkan pengukuran Berat Badan sesuai Tinggi Badan (BB/TB) menjadi lima kategori status gizi yaitu Gizi Buruk, Gizi Kurang, Gizi Baik, Resiko Gizi Lebih, Gizi Lebih. Dari kelima kategori yang dihasilkan tersebut harus dicocokkan dengan menggunakan table pertumbuhan yang berlaku untuk suatu masyarakat, sehingga dapat mengevaluasi dan melihat perkembangan fisik anak.

Hasil penelitian menunjukkan kinerja *Algoritma KNN* dengan akurasi 91,26% lebih baik dari hasil kinerja *Algoritma C4.5* dengan akurasi 91,12%.

Kata kunci : *Data Mining, Status Gizi Anak, Klasifikasi, Decision Tree, Algoritma C4.5, Algoritma K-Nearest Neighbor (KNN)*.

ABSTRACT

The health of children aged 0-60 months must be optimal to shape a golden generation in their time, and nutrition is a crucial intake that influences a child's growth and development. The nutritional needs of children, especially toddlers, are highly susceptible to changes in weight. To assess nutritional status, toddlers undergo anthropometric examinations, including periodic measurements of height and weight. The examination results are manually recorded on a Growth Monitoring Card (KMS). This method is considered less accurate and incomplete because nutritional status assessment is solely based on weight-for-age (W/A) of toddlers.

Therefore, data mining algorithms are employed to classify the nutritional status of toddlers. The study aims to determine which algorithm, either C.45 or K-Nearest Neighbour (KNN), is the most effective or accurate in classifying the nutritional status of toddlers. RapidMiner version 10.3 is used as a tool, considering four factors: 1) Gender, 2) Age, 3) Weight-for-Age (W/A) measurement, and 4) Height-for-Age (H/A) measurement. The results are then transformed into Weight-for-Height (W/H) measurements, categorizing nutritional status into five categories: Malnutrition, Undernutrition, Normal Nutrition, Overnutrition Risk, and Overnutrition. These categories are then compared with growth charts applicable to a specific community to evaluate and monitor the physical development of children.

The research results indicate that the performance of the KNN algorithm, with an accuracy of 91.26%, is better than the C45 algorithm, which achieved an accuracy of 91.12%.

Keywords: Data Mining, Child Nutritional Status, Classification, Decision Tree, C4.5 Algorithm, K-Nearest Neighbor Algorithm (KNN)