

**IMPLEMENTASI *MACHINE LEARNING* ALGORITMA C4.5  
DALAM PENGKLASIFIKASIAN PASIEN YANG  
MENGALAMI PENYAKIT DIABETES MELLITUS TIPE 2**

**SKRIPSI**

*Diajukan sebagai salah satu persyaratan untuk memperoleh gelar  
Sarjana Sains*



**Oleh:  
DYAH AYU SEKAR KINASHIH PURWANINGRUM  
NIM. 19030096**

**PROGRAM STUDI MATEMATIKA  
DEPARTEMEN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2023**

**IMPLEMENTASI *MACHINE LEARNING* ALGORITMA C4.5  
DALAM PENGKLASIFIKASIAN PASIEN YANG  
MENGALAMI PENYAKIT DIABETES MELLITUS TIPE 2**

**SKRIPSI**

*Diajukan sebagai salah satu persyaratan untuk memperoleh gelar  
Sarjana Sains*



**Oleh:**

**DYAH AYU SEKAR KINASHIH PURWANINGRUM  
NIM. 19030096**

**PROGRAM STUDI MATEMATIKA  
DEPARTEMEN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2023**

**PERSETUJUAN SKRIPSI**

**IMPLEMENTASI *MACHINE LEARNING* ALGORITMA C4.5 DALAM  
PENGKLASIFIKASIAN PASIEN YANG MENGALAMI  
PENYAKIT DIABETES MELLITUS TIPE 2**

Nama : Dyah Ayu Sekar Kinasih Purwaningrum  
NIM : 19030096  
Program Studi : Matematika  
Departemen : Matematika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 21 Agustus 2023

Disetujui oleh,

Pembimbing



Dina Agustina, S.Pd., M.Sc

NIP. 199208312020122015

**PENGESAHAN LULUS UJIAN SKRIPSI**

Nama : Dyah Ayu Sekar Kinasih Purwaningrum  
NIM : 19030096  
Program Studi : Matematika  
Departemen : Matematika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

**IMPLEMENTASI *MACHINE LEARNING* ALGORITMA C4.5 DALAM  
PENGKLASIFIKASIAN PASIEN YANG MENGALAMI  
PENYAKIT DIABETES MELLITUS TIPE 2**

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi  
Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Negeri Padang

Padang, 21 Agustus 2023

Tim Penguji

	Nama
Ketua	: Dina Agustina, S.Pd., M.Sc
Anggota	: Dra. Helma, M.Si
Anggota	: Drs. Yusmet Rizal, M.Si

Tanda Tangan




## SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertanda tangan di bawah ini:

Nama : Dyah Ayu Sekar Kinasih Purwaningrum  
NIM : 19030096  
Program Studi : Matematika  
Departemen : Matematika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan, bahwa skripsi saya dengan judul "**Implementasi *Machine Learning* Algoritma C4.5 dalam Pengklasifikasian Pasien yang Mengalami Diabetes Mellitus Tipe 2**" adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan. Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikianlah pernyataan ini saya buat dengan penuh kesadaran dan rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Padang, 21 Agustus 2023

Diketahui oleh,  
Kepala Departemen Matematika,



Dr. Suherman, S.Pd., M.Si  
NIM. 196808301999031002

Saya yang menyatakan,

  


Dyah Ayu Sekar Kinasih  
NIM. 19030096

# Implementasi *Machine Learning* Algoritma C4.5 Dalam Pengklasifikasian Pasien Yang Mengalami Penyakit Diabetes Mellitus Tipe 2

Dyah Ayu Sekar Kinasih Purwaningrum

## ABSTRAK

Pola hidup sehat dapat mengurangi risiko berbagai penyakit, tetapi banyak dari masyarakat Indonesia yang masih mengabaikannya. Pengabaian tersebut dapat mengundang penyakit diabetes mellitus yang sudah menjangkiti 643 juta orang di dunia. Sehingga, diperlukan diagnosa dini yang akurat sebagai bentuk pencegahan diabetes. Dengan menggunakan *machine learning* algoritma C4.5, penelitian ini akan membentuk model yang mampu mengklasifikasikan apakah seseorang dikatakan diabetes atau tidak berdasarkan faktor penyebab diabetes.

Penelitian ini merupakan penelitian terapan yang diawali dengan studi literatur terkait teori-teori yang relevan dengan permasalahan penelitian, kemudian dilanjutkan dengan pengumpulan data. Data yang digunakan berupa data rekam medis pasien diabetes dan tidak diabetes Rumah Sakit Umum Pusat (RSUP) Dr. M. Djamil Padang. Data yang telah terkumpul selanjutnya diproses menggunakan algoritma C4.5.

Model decision tree yang dihasilkan terdiri dari 33 nodes dengan atribut glukosa sebagai root *node*-nya. Pada masing-masing *node* terdapat keterangan atribut, *entropy*, sampel, nilai, dan kelas. Dari evaluasi kinerja model dihasilkan nilai *recall* 94%.

Kata Kunci: Algoritma C4.5, Diabetes Mellitus, Klasifikasi, *Machine Learning*

# **Implementation Of Machine Learning Algorithm C4.5 In Classification Of Patient With Type 2 Diabetes Mellitus**

**Dyah Ayu Sekar Kinasih Purwaningrum**

## **ABSTRACT**

A healthy lifestyle can minimize the risk of numerous diseases; however, many Indonesians still disregard it. The neglect of a healthy lifestyle can lead to diabetes mellitus which has infected 643 million people worldwide. Hence, early diagnosis is important in preventing diabetes. Using the machine learning C4.5 algorithm, this study will form a model that can classify whether a person has diabetes or not based on diabetes risk factors.

This is an applied study that begins with a literature review on theories relevant to the research topic and then continues to collect data. The data used is from the medical records of both diabetic and non-diabetic patients at the Central Hospital Dr. M. Djamil Padang. The data collected is then processed using the C4.5 algorithm.

The decision tree model developed has 33 nodes, with the attribute glucose as the root node. Each node contains an attribute, entropy, sample, value, and class. The model's performance evaluation generates a 94% recall.

Keywords: C4.5 Algorithm, Classification, Diabetes Mellitus, Machine Learning

## KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Puji syukur peneliti panjatkan kepada Allah SWT yang telah memberikan keridhaan-Nya kepada peneliti sehingga peneliti dapat menyelesaikan tugas akhir dengan judul “Implementasi *Machine Learning* Algoritma C4.5 dalam Pengklasifikasian Pasien yang Mengalami Diabetes Tipe 2”. Tugas akhir ini merupakan salah satu syarat dalam memperoleh gelar Sarjana Sains pada Program Studi Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Negeri Padang (UNP).

Dalam penyelesaian tugas akhir ini, peneliti mendapat banyak bimbingan, dukungan, nasehat, serta bantuan dari berbagai pihak sehingga pada akhirnya pengerjaan tugas akhir ini dapat terselesaikan. Ucapan terima kasih peneliti sampaikan kepada:

1. Ibu Dina Agustina, S.Pd., M.Sc, sebagai Dosen Pembimbing sekaligus Penasehat Akademik
2. Ibu Dr. Devni Prima Sari, S.Si., M.Sc, Ibu Dra. Helma, M.Si, dan Bapak Drs. Yusmet Rizal, M.Si, sebagai Dosen Penguji
3. Bapak Defri Ahmad, S.Pd., M.Si yang menjabat sebagai Ketua Program Studi Matematika UNP, Ketua Departemen Matematika UNP, dan Dosen Penguji
4. Bapak dan Ibu Dosen, Staf Pengajar, dan Karyawan Departemen Matematika FMIPA UNP



5. Pihak RSUP Dr. M. Djamil Padang, khususnya bagian rekam medis yang telah memberikan izin dan bantuan dalam pengumpulan data yang dibutuhkan pada penelitian ini
6. Pihak lain yang telah membantu proses pengerjaan tugas akhir ini

Semoga Allah SWT membalas semua perbuatan baik yang telah dilakukan. Akhir kata, peneliti menyadari bahwa tugas akhir ini masih jauh dari kata sempurna. Oleh karena itu peneliti akan sangat terbuka terhadap kritik dan saran yang bersifat membangun dari semua pihak demi kesempurnaan tugas akhir ini. Peneliti juga berharap semoga tugas akhir ini akan bermanfaat bagi pembacanya.

Padang, 22 Juli 2023

Dyah Ayu Sekar Kinasih

## UCAPAN TERIMA KASIH

### *The Acknowledgement*

*a personal thank you to:*

Ibuk, Papa, Adek. *Every time I was ready to quit, or something bothered me, I am just gonna put on a call for one of you and then I healed. You guys are always healing my heart that you didn't break, and I am forever grateful. This final project stands as a testament to your unconditional love and encouragement.*

Rida Purnama Sari. *You are one of my answered prayers. Truly. Thank you for always being there for me, I owe you a lot.*

Nurul Khasanah. *In the scenes of Phineas and Ferb, you're the Stacy for me <3*

Nurul Aulia, Chika Nuragustirani, and Risti Widyastiti. *Who opened both their room and heart to me when I first arrived in the city.*

*Beloved me, Dyah Ayu Sekar Kinasih. Thank you for willing to admit your failures, your shortcomings, and your lack of judgement at times. I sincerely hope it pays off-and I think it will. Just know that I have nothing but gratitude for the woman you are and will continue to be.*

### **A nice quote**

*Allow yourself to be proud of yourself and all the progress you have made.*

## DAFTAR ISI

ABSTRAK .....	i
KATA PENGANTAR .....	iii
DAFTAR ISI .....	vi
DAFTAR GAMBAR .....	vii
DAFTAR TABEL .....	viii
DAFTAR LAMPIRAN .....	ix
BAB I PENDAHULUAN .....	1
A. Latar Belakang Penelitian .....	1
B. Rumusan Masalah .....	3
C. Batasan Masalah .....	3
D. Tujuan Penelitian .....	4
E. Manfaat Penelitian .....	4
BAB II LANDASAN TEORI .....	5
A. Tinjauan Pustaka .....	5
B. Diabetes Mellitus .....	6
C. <i>Machine Learning</i> .....	9
D. <i>Decision Tree</i> .....	10
E. Algoritma C4.5 .....	11
F. Evaluasi Kinerja Model .....	15
BAB III METODOLOGI PENELITIAN .....	20
A. Jenis Penelitian .....	20
B. Jenis dan Sumber Data .....	20
C. Kerangka Kerja Penelitian .....	20
BAB IV HASIL DAN PEMBAHASAN .....	23
A. Deskripsi Data .....	23
B. Pembahasan .....	27
BAB V PENUTUP .....	36
A. Kesimpulan .....	36
B. Saran .....	36
DAFTAR PUSTAKA .....	38
LAMPIRAN .....	41

## DAFTAR GAMBAR

Gambar 1. Struktur Decision Tree .....	11
Gambar 2. Diagram Alir Algoritma C4.5 .....	13
Gambar 3 Contoh Kurva ROC dan AUC.....	18
Gambar 4. Kerangka Kerja Penelitian .....	21
Gambar 5. Dataset yang Digunakan.....	23
Gambar 6. Distribusi Kelas Dataset Diabetes .....	25
Gambar 7. Distribusi Kelas Dataset Terhadap Atribut Numerikal .....	26
Gambar 8. Distribusi Kelas Dataset Terhadap Atribut Kategorikal .....	26
Gambar 9. Bentuk <i>Decision Tree</i> pada Proses Awal .....	29
Gambar 10. Bentuk <i>Decision Tree</i> dengan Tiga <i>Node</i> .....	30
Gambar 11. Bentuk <i>Decision Tree</i> dengan Lima <i>Node</i> .....	32
Gambar 12. <i>Decision Tree</i> Data Diabetes.....	32
Gambar 13. <i>Confusion Matrix</i> yang Dihasilkan dari (a) Data <i>Testing</i> , dan (b) Data <i>Training</i> .....	33
Gambar 14. Kurva ROC.....	35

## DAFTAR TABEL

Tabel 1. Kategori BMI .....	8
Tabel 3. <i>Confusion Matrix</i> .....	16
Tabel 4. Deskripsi Atribut Dataset.....	23
Tabel 5. Deskripsi Statistik Atribut Numerikal.....	24
Tabel 6. Deskripsi Statistik Atribut Kategorikal.....	25
Tabel 7. Konversi Atribut Kategorikal.....	27
Tabel 8. Pemilihan Atribut <i>Root Node</i> .....	28
Tabel 9. Pemilihan Atribut <i>Splitting</i> Kondisi <i>True</i> .....	29
Tabel 10. Pemilihan Atribut <i>Node 1.2</i> .....	31
Tabel 11. Hasil Evaluasi Kinerja Model.....	34

## DAFTAR LAMPIRAN

Lampiran 1. Dataset .....	41
Lampiran 2. <i>Threshold</i> Atribut Numerikal .....	54
Lampiran 3. <i>Code Python</i> .....	61
Lampiran 4. Model <i>Decision Tree</i> .....	65

# **BAB I**

## **PENDAHULUAN**

### **A. Latar Belakang Penelitian**

Penerapan pola hidup sehat dapat mengurangi risiko berbagai penyakit, tetapi banyak dari masyarakat Indonesia yang masih mengabaikannya. Hal ini dibuktikan oleh survey AIA Grup di 15 negara Asia Pasifik yang menempatkan Indonesia sebagai peringkat 11 dalam penerapan pola hidup sehat. Akibatnya, masyarakat Indonesia rentan mengalami berbagai penyakit, salah satunya penyakit diabetes mellitus.

Diabetes mellitus atau yang biasanya hanya disebut dengan diabetes merupakan penyakit yang disebabkan karena hiperglikemia, yaitu kondisi menumpuknya kadar glukosa dalam darah. Walaupun demikian, diyakini bahwa diabetes tipe 2 juga disebabkan oleh beberapa faktor seperti *body mass index* berlebih, bertambahnya umur, dan riwayat keluarga (IDF Diabetes Atlas 10th Edition, 2021). Apabila penumpukan terjadi akibat penyerangan sistem imun terhadap sel pankreas penghasil insulin maka disebut diabetes tipe 1, sedangkan apabila terjadinya akibat produksi insulin yang kurang atau sel tubuh yang tidak dapat menggunakan hormon insulin secara efektif maka disebut diabetes tipe 2 (Maniruzzaman et al., 2017). Jumlah kasus kedua tipe diabetes terus meningkat, tetapi diabetes tipe 2 memiliki rasio yang lebih besar yaitu 90% dari jumlah penderita diabetes secara keseluruhan (IDF Diabetes Atlas 10th Edition, 2021).

Pada 2021, sedikitnya 643 juta orang di dunia menderita diabetes. Indonesia terletak pada urutan kelima sebagai negara dengan jumlah penderita

diabetes tertinggi di dunia, mencapai 19,5 juta dan diprediksi akan terus meningkat hingga 28,6 juta di tahun 2045 (IDF Diabetes Atlas 10th Edition, 2021). Tingginya angka statistik ini tentu mengharuskan kita melakukan berbagai upaya guna mencegah peningkatan angka pasien diabetes, salah satunya dengan memanfaatkan data hasil uji lab pasien (rekam medis) (FS et al., 2010).

Dalam bidang kesehatan, terdapat banyak data rekam medis yang akan berakhir menjadi data yang tidak berguna apabila tidak dimanfaatkan dengan baik (Ente et al., 2019). Untuk itu diperlukanlah sebuah proses pemanfaatan data rekam medis guna mendapatkan informasi yang akan berguna dalam upaya penekanan angka pasien diabetes. Salah satu metode yang dapat digunakan untuk proses ini adalah *machine learning*.

Berdasarkan penelitian terdahulu, pendekatan *machine learning* meningkatkan akurasi pada prediksi faktor penyebab penyakit dibandingkan dengan pendekatan konvensional (Selya & Anshutz, 2018). Mercaldo (2017) menggunakan *machine learning decision tree* untuk memprediksi diabetes berdasarkan kadar glukosa, *body mass index*, umur, dan atribut lainnya. Sementara itu, Azrar (2018) menggunakan algoritma C4.5 dalam mengklasifikasikan perempuan yang mengalami diabetes. Dengan menggunakan Pima Indian Diabetes Dataset, didapatkan nilai akurasi tertinggi adalah dengan menggunakan algoritma C4.5 yaitu 75,65%. Ente (2019) juga telah membuktikan bahwa algoritma C4.5 memberikan hasil yang lebih baik daripada metode *machine learning* lainnya dalam mengklasifikasikan pasien diabetes.



Pada dasarnya algoritma C4.5 memiliki konsep mengubah data kompleks menjadi model berbentuk pohon dan aturan-aturan keputusannya (*decision tree*) (Ruggieri, 2002). Tidak seperti algoritma klasifikasi lainnya, algoritma C4.5 mampu menangani atribut kategorikal maupun numerikal (Quinlan, 1993), seperti atribut yang akan digunakan pada penelitian ini.

Sehingga berdasarkan informasi yang telah dipaparkan, akan dilakukan penelitian berjudul “Implementasi *Machine learning* Algoritma C4.5 dalam Pengklasifikasian Pasien yang Mengalami Penyakit Diabetes Tipe 2”.

## **B. Rumusan Masalah**

Berdasarkan latar belakang masalah yang telah dipaparkan, maka rumusan masalah pada penelitian ini adalah “Bagaimana Penerapan *Machine Learning* Algoritma C4.5 dalam Mengklasifikasikan Pasien Diabetes Tipe 2”.

## **C. Batasan Masalah**

Untuk membatasi masalah penelitian agar tidak terjadi pelebaran pokok masalah, ditetapkanlah batasan masalah sebagai berikut:

1. Data yang diolah merupakan data rekam medis pasien diabetes dan tidak diabetes RSUP Dr. M. Djamil Padang.
2. Penyakit diabetes yang dimaksud adalah diabetes tipe 2.
3. Algoritma yang digunakan adalah algoritma C4.5 beserta evaluasi kinerja model yang dihasilkan.
4. Bahasa pemrograman yang digunakan adalah *python*, dengan *Google Collaboratory* sebagai *Integrated Development Environment* (IDE) – nya.

#### **D. Tujuan Penelitian**

Dari permasalahan yang disebutkan maka tujuan dari penelitian ini adalah:

1. Membentuk model *decision tree* klasifikasi pasien diabetes tipe 2 menggunakan algoritma C4.5
2. Melakukan evaluasi kinerja model berdasarkan *confusion matrix* yang terbentuk

#### **E. Manfaat Penelitian**

Berdasarkan tujuan penelitian maka manfaat yang dihasilkan dari penelitian ini adalah:

1. Menambah pengetahuan, wawasan, dan pemahaman peneliti dan pembaca mengenai penerapan algoritma C4.5 dalam kasus klasifikasi menggunakan *machine learning*.
2. Membantu masyarakat khususnya tenaga kesehatan untuk mendiagnosa dini apakah seseorang dikatakan diabetes atau tidak.
3. Acuan dan bahan referensi untuk pembaca dan peneliti yang akan melakukan penelitian selanjutnya.