



**IMPLEMENTASI METODE *BAGGING* DAN *ADABOOST* PADA  
*ALGORITMA C5.0* UNTUK MENDIAGNOSA PENYAKIT DIABETES  
*MELITUS***

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh**

**NURMOLIKA**

**NIM 1911016220013**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU  
JUNI 2023**



**IMPLEMENTASI METODE *BAGGING* DAN *ADABOOST* PADA  
*ALGORITMA C5.0* UNTUK MENDIAGNOSA PENYAKIT DIABETES  
*MELITUS***

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh**

**NURMOLIKA  
NIM 1911016220013**

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

**JUNI 2023**

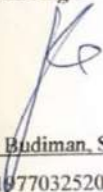
**SKRIPSI**  
**IMPLEMENTASI METODE *BAGGING* DAN *ADABOOST* PADA**  
**ALGORITMA *C5.0* UNTUK MENDIAGNOSA PENYAKIT DIABETES**  
***MELITUS***

Oleh:  
**NURMOLIKA**  
**NIM. 1911016220013**


Telah dipertahankan di depan Dosen Penguji pada tanggal 27 Juni 2023.

Susunan Dosen Penguji:

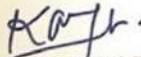
**Pembimbing I**

  
Irwan Budiman, S.T., M.Kom  
NIP. 197703252008121001


**Dosen Penguji I**

  
Fatma Indriani, S.T., M.I.T., Ph.D  
NIP. 198404202008122004

**Pembimbing II**

  
Dwi Kartini, S.Kom., M.Kom  
NIP. 198704212012122003

**Dosen Penguji II**

  
Andi Firmadi, S.Si., M.T  
NIP. 197307252008011006

Banjarbaru, 27 Juni 2023

**Coordinator Program Studi Ilmu Komputer**



Irwan Budiman, S.T., M.Kom  
NIP. 197703252008121001

## 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, .. Juni 2023

Yang Menyatakan,



Nurmolika

NIM.1911016220013

## ABSTRAK

### **IMPLEMENTASI METODE *BAGGING* DAN *ADABOOST* PADA ALGORITMA *C5.0* UNTUK MENDIAGNOSA PENYAKIT DIABETES MELITUS**

(Oleh : Nurmolika; Pembimbing: Irwan Budiman, S.T., M.Kom dan Dwi Kartini, S.Kom., M.Kom.; 2023; 95 halaman)

Diabetes Melitus atau kencing manis adalah gangguan kesehatan dimana jumlah atau konsentrasi glukosa atau gula dalam darah melebihi atau tidak sesuai dengan kondisi normal. Dataset yang digunakan yaitu dataset diabetes diperoleh dari website Kaggle datasets. Data diabetes melitus ini terdiri dari 768 data. Dimana data pasien yang negatif diabetes mellitus sebanyak 500 data dan pasien yang positif diabetes mellitus sebanyak 268 data. Data ini terdiri dari 9 atribut. Data yang ada dianalisis menggunakan algoritma *C5.0*, algoritma *C5.0* dengan metode *bagging*, dan algoritma *C5.0* dengan metode *adaboost*. Penggunaan metode *bagging* dan *adaboost* untuk menangani data tidak seimbang yang terdapat pada data diabetes melitus. Hasil klasifikasi berupa *accuracy* dengan menggunakan data training 80% dan data testing 20%. Nilai *accuracy* klasifikasi menggunakan algoritma *C5.0* sebesar 77,27%, algoritma *C5.0* dengan metode *bagging* sebesar 75,97%, dan algoritma *C5.0* dengan metode *adaboost* sebesar 75,97%. Dalam penelitian ini, algoritma *C5.0* memiliki nilai *accuracy* yang paling baik.

**Kata kunci:** Diabetes Melitus, Klasifikasi, *C5.0*, *Bagging*, *Adaboost*.

## ***ABSTRACT***

### ***IMPLEMENTATION OF BAGGING AND ADABOOST METHODS ON THE C5.0 ALGORITHM TO DIAGNOSE DIABETES MELLITUS***

(By : Nurmolika; Supervisors: Irwan Budiman, S.T., M.Kom and Dwi Kartini, S.Kom., M.Kom.; 2023; 95 pages)

Diabetes mellitus or diabetes is a health disorder where the amount or concentration of glucose or sugar in the blood exceeds or does not match normal conditions. The dataset used is a diabetes dataset obtained from the Kaggle datasets website. This diabetes mellitus data consists of 768 data. Where the data of patients who are negative for diabetes mellitus as many as 500 data and patients who are positive for diabetes mellitus as many as 268 data. This data consists of 9 attributes. The existing data was analyzed using the C5.0 algorithm, the C5.0 algorithm with the bagging method, and the C5.0 algorithm with the adaboost method. The use of bagging methods and adaboost to handle unbalanced data contained in diabetes mellitus data. The classification results are in the form of accuracy using 80% training data and 20% testing data. The accuracy value classification using the C5.0 algorithm is 77.27%, the C5.0 algorithm with the bagging method is 75.97%, and the C5.0 algorithm with the adaboost method is 75.97%. In this study, the C5.0 algorithm has the best accuracy value.

***Keywords:*** Diabetes mellitus, Classification, C5.0, *Bagging*, *Adaboost*.

## PRAKATA

Puji syukur penulis panjatkan ke Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “IMPLEMENTASI METODE *BAGGING* DAN *ADABOOST PADA ALGORITMA C5.0 UNTUK MENDIAGNOSA PENYAKIT DIABETES MELITUS*” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

1. Keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini
2. Bapak Irwan Budiman, S.T., M.Kom selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Ibu Dwi Kartini, S. Kom., M. Kom selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2019 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
7. Muhammad Rahmat Dani, terima kasih atas dukungan, semangat, serta telah menjadi tempat berkeluh kesah, selalu ada dalam suka maupun duka selama proses penyusunan skripsi ini

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 23 Juni 2023

A handwritten signature in black ink, appearing to be 'Nurmolika', written in a cursive style.

Nurmolika



## DAFTAR ISI

	<b>Halaman</b>
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	<b>Kesalahan! Bookmark tidak ditentukan.</b>
HALAMAN PERNYATAAN .....	iii
ABSTRAK .....	iv
<i>ABSTRACT</i> .....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xiv
DAFTAR LAMPIRAN.....	xvi
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian .....	3
1.4 Manfaat Penelitian .....	3
1.5 Batasan Masalah .....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Terdahulu .....	5
2.2 Landasan Teori .....	9
2.2.1 Diabetes Melitus.....	9
2.2.2 Data Mining .....	9
2.2.3 Decision Tree .....	11
2.2.4 C5.0.....	14
2.2.5 Teknik Ensemble.....	17
2.2.6 Metode <i>Bagging</i> .....	17

2.2.7	Metode Adaboost .....	18
2.2.8	Imputasi <i>Mean</i> .....	20
2.2.9	Split Data.....	20
2.2.10	Confusion Matrix .....	21
BAB III METODE PENELITIAN.....		23
3.1	Bahan Penelitian .....	23
3.2	Alat Penelitian .....	23
3.3	Variabel Penelitian.....	25
3.4	Prosedur Penelitian .....	25
BAB IV HASIL DAN PEMBAHASAN .....		28
4.1	Hasil.....	28
4.1.1	Pengumpulan Dataset.....	28
4.1.2	<i>Preprocessing Data</i> .....	29
4.1.3	Data Mining .....	34
4.1.3.1	Algoritma C5.0.....	34
4.1.3.2	Algoritma C5.0 dengan Metode <i>Bagging</i> .....	54
4.1.3.3	Algoritma C5.0 dengan Metode Adaboost.....	78
4.1.4	Evaluasi.....	88
4.1.4.1	Evaluasi Algoritma <i>C5.0</i> .....	88
4.1.4.2	Evaluasi Algoritma C5.0 dengan Metode <i>Bagging</i> .....	89
4.1.4.3	Evaluasi Algoritma C5.0 dengan Metode <i>Adaboost</i> ....	89
4.2	Pembahasan .....	90
4.2.1	Algoritma <i>C5.0</i> .....	91
4.2.2	Algoritma <i>C5.0</i> dengan Metode <i>Bagging</i> .....	92
4.2.3	Algoritma <i>C5.0</i> dengan Metode <i>Adaboost</i> .....	92

4.2.4 Akurasi Kedalaman Pohon Keputusan .....	95
BAB V PENUTUP.....	96
5.1 Kesimpulan.....	96
5.2 Saran .....	96
DAFTAR PUSTAKA .....	97
LAMPIRAN.....	101

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
Tabel 1. Keaslian Penelitian.....	6
Tabel 2. Perancangan Penelitian .....	8
Tabel 3. <i>Confusion Matrix</i> (Rais, 2019) .....	21
Tabel 4. Keterangan Nilai Atribut Pada Dataset Diabetes .....	23
Tabel 5. Sebaran Kelas.....	23
Tabel 6. Dataset Diabetes Melitus .....	29
Tabel 7. Rincian atribut ada nilai 0 pada Dataset Diabetes.....	29
Tabel 8. Rincian <i>Missing value</i> berseta nilai <i>mean</i> . .....	30
Tabel 9. Dataset Diabetes setelah di <i>Cleaning</i> .....	30
Tabel 10. Hasil <i>Transformation</i> data .....	31
Tabel 11. Dataset Diabetes setelah di <i>Transformation</i> .....	32
Tabel 12. Dataset Training .....	33
Tabel 13. Dataset Testing.....	33
Tabel 14. Perhitungan node 1.....	36
Tabel 15. Perhitungan node 1.1.....	37
Tabel 16. Perhitungan Node 1.1.1.....	39
Tabel 17. Perhitungan Node 1.1.2.....	41
Tabel 18. Perhitungan Node 1.1.3.....	43
Tabel 19. Perhitungan Node 1.2.....	44
Tabel 20. Perhitungan Node 1.2.1.....	46
Tabel 21. Perhitungan Node 1.2.2.....	48
Tabel 22. Perhitungan Node 1.2.3.....	49
Tabel 23. Hasil Klasifikasi hasil <i>Data Testing</i> menggunakan algoritma C5.0 ..	53
Tabel 24. Data pada <i>Bagging</i> 1 .....	54
Tabel 25. Data pada <i>Bagging</i> 2 .....	54
Tabel 26. Data pada <i>Bagging</i> 3 .....	55
Tabel 27. Data pada <i>Bagging</i> 4 .....	55
Tabel 28. Data pada <i>Bagging</i> 5 .....	56

Tabel 29. Data pada <i>Bagging</i> 6 .....	56
Tabel 30. Data pada <i>Bagging</i> 7 .....	57
Tabel 31. Data pada <i>Bagging</i> 8 .....	57
Tabel 32. Data pada <i>Bagging</i> 9 .....	57
Tabel 33. Data pada <i>Bagging</i> 10 .....	58
Tabel 34. Hasil Klasifikasi Data <i>Testing</i> menggunakan Algoritma <i>C5.0</i> dengan Metode <i>Bagging</i> .....	77
Tabel 35. Data pada Adaboost 1 .....	78
Tabel 36. Hasil Klasifikasi Data <i>Training</i> dengan menggunakan Algoritma <i>C5.0</i> pada <i>Adaboost</i> 1 .....	79
Tabel 37. Pembaruan Bobot Setiap Data pada Data <i>Training</i> setelah <i>Adaboost</i> 1 .....	80
Tabel 38. Data pada Adaboost 2 .....	81
Tabel 39. Hasil Klasifikasi Data <i>Training</i> dengan menggunakan Algoritma <i>C5.0</i> pada <i>Adaboost</i> 2 .....	82
Tabel 40. Pembaruan Bobot Setiap Data pada Data <i>Training</i> setelah <i>Adaboost</i> 2 .....	83
Tabel 41. Data pada Adaboost 3 .....	84
Tabel 42. Hasil Klasifikasi Data <i>Training</i> dengan menggunakan Algoritma <i>C5.0</i> pada <i>Adaboost</i> 3 .....	85
Tabel 43. Pembaruan Bobot Setiap Data pada Data <i>Training</i> setelah <i>Adaboost</i> 3 .....	86
Tabel 44. Hasil Klasifikasi Data <i>Testing</i> menggunakan Algoritma <i>C5.0</i> pada Setiap <i>Adaboost</i> . Perhitungan $H(x)$ , dan Klasifikasi Akhir .....	87
Tabel 45. <i>Confusion Matrix</i> Hasil Klasifikasi Menggunakan Algoritma <i>C5.0</i> .....	88
Tabel 46. <i>Confusion Matrix</i> Hasil Klasifikasi Menggunakan Algoritma <i>C5.0</i> dengan Metode <i>Bagging</i> .....	89
Tabel 47. <i>Confusion Matrix</i> Hasil Klasifikasi Menggunakan Algoritma <i>C5.0</i> dengan Metode <i>Adaboost</i> .....	89
Tabel 48. Hasil Klasifikasi Data Menggunakan Algoritma <i>C5.0</i> .....	92

Tabel 49. Hasil Klasifikasi Data Menggunakan Algoritma <i>C5.0</i> dengan Metode <i>Bagging</i> .....	92
Tabel 50. Hasil Klasifikasi Data Menggunakan Algoritma <i>C5.0</i> dengan Metode <i>Adaboost</i> .....	93

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
Gambar 1. Tahapan Data Mining.....	10
Gambar 2. Pohon Keputusan.....	12
Gambar 3. Syarat pengujian fitur biner.....	13
Gambar 4. Syarat Pengujian Fitur Bertipe Nominal .....	13
Gambar 5. Syarat Pengujian Fitur Bertipe Ordinal.....	14
Gambar 6 Syarat pengujian fitur bertipe numerik .....	14
Gambar 7. Contoh pohon keputusan.....	16
Gambar 8. Skema model Bagging .....	18
Gambar 9. Alur Penelitian.....	25
Gambar 10. Pohon Keputusan Node 1 .....	37
Gambar 11. Pohon Keputusan Node 1.1 .....	39
Gambar 12. Pohon Keputusan 1.1.1.....	41
Gambar 13. Pohon Keputusan Node 1.1.2.....	42
Gambar 14. Pohon Keputusan Node 1.1.3.....	44
Gambar 15. Pohon Keputusan Node 1.2.....	46
Gambar 16. Pohon Kepohonan Node 1.2.1.....	47
Gambar 17. Pohon Keputusan Node 1.2.2.....	49
Gambar 18. Pohon Keputusan Node 1.2.3 .....	51
Gambar 19. Pohon Keputusan Diabetes Melitus .....	52
Gambar 20. Pohon Keputusan <i>Bagging</i> 1 .....	60
Gambar 21. Pohon Keputusan <i>Bagging</i> 2.....	62
Gambar 22. Pohon Keputusan <i>Bagging</i> 3 .....	64
Gambar 23. Pohon Keputusan <i>Bagging</i> 4.....	66
Gambar 24. Pohon Keputusan <i>Bagging</i> 5 .....	67
Gambar 25. Pohon Keputusan <i>Bagging</i> 6.....	69
Gambar 26. Pohon Keputusan <i>Bagging</i> 7.....	71
Gambar 27. Pohon Keputusan <i>Bagging</i> 8.....	73
Gambar 28. Pohon Keputusan <i>Bagging</i> 9.....	75

Gambar 29. Pohon Keputusan <i>Bagging</i> 10 .....	77
Gambar 30. Pohon Keputusan <i>Adaboost</i> 1 .....	79
Gambar 31. Pohon Keputusan <i>Adaboost</i> 2 .....	82
Gambar 32. Pohon Keputusan <i>Adaboost</i> 3 .....	85
Gambar 33. Presentase Evaluasi Menggunakan Confusion Matrix.....	93
Gambar 34. Akurasi Kedalaman Pohon.....	95



## DAFTAR LAMPIRAN

### **Lampiran**

Lampiran 1 Dataset *Original*

Lampiran 2 Dataset *Cleaning*

Lampiran 3 Data *Transformation*

Lampiran 4 Data *Training*

Lampiran 5 Data *Testing*

Lampiran 6 Data *Bagging*

Lampiran 7 Data *Adaboost*

Lampiran 8 Hasil Klasifikasi Algoritma *C5.0*

Lampiran 9 Hasil Klasifikasi Algoritma *C5.0* dengan Metode *Bagging*

Lampiran 10 Hasil Klasifikasi Algoritma *C5.0* dengan Metode *Adaboost*

Lampiran 11 Gambar Pohon Keputusan *Bagging*

Lampiran 12 Gambar Pohon Keputusan *Adaboost*