

**SISTEM KLASIFIKASI FAKTOR RESIKO DIABETES MELLITUS
MENGUNAKAN METODE DECISION TREE C4.5**

SKRIPSI



disusun oleh

Agus Triyanto

17.11.0955

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023**

**SISTEM KLASIFIKASI FAKTOR RESIKO DIABETES MELLITUS
MENGUNAKAN METODE DECISION TREE C4.5**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai Gelar Sarjana
pada Program Studi Informatika



disusun oleh
Agus Triyanto
17.11.0955

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023**

PERSETUJUAN

SKRIPSI

**SISTEM KLASIFIKASI FAKTOR RESIKO DIABETES MELLITUS
MENGUNAKAN METODE DECISION TREE C4.5**

yang dipersiapkan dan disusun oleh

Agus Triyanto

17.11.0955

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Maret 2023

Dosen Pembimbing,



Erni Seniwati, S.Kom., M.Cs.

NIK. 190302231

PENGESAHAN

SKRIPSI

**SISTEM KLASIFIKASI FAKTOR RESIKO DIABETES MELLITUS
MENGUNAKAN METODE DECISION TREE C4.5**

yang dipersiapkan dan disusun oleh

Agus Triyanto

17.11.0955

telah dipertahankan didepan dewan penguji
pada tanggal 21 maret 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ferry Wahyu Wibowo, S.Si, M.Cs.

NIK. 190302235

Ria Andriani, M.Kom.

NIK. 190302458

Erni Seniwati, S.Kom, M.Cs.

NIK. 190302231



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar sarjana komputer
Tanggal 29 Maret 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, M.Kom.

NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 29 Maret 2023



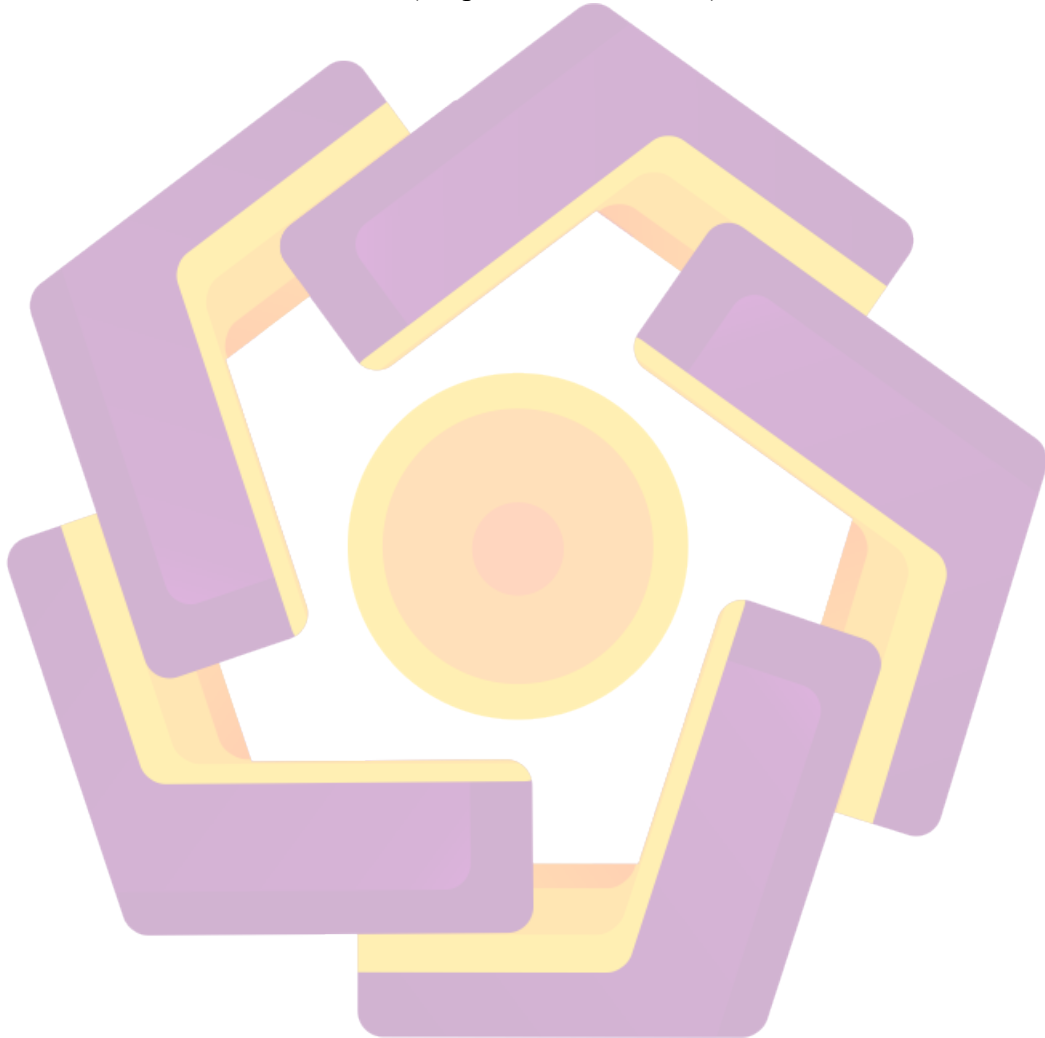
Agus Triyanto

17.11.0955

MOTTO

“Jangan pergi mengikuti kemana jalan akan berujung. Buat jalanmu sendiri dan tinggalkanlah jejak.”

(Ralph Waldo Emerson)



PERSEMBAHAN

Dengan rasa syukur yang mendalam, dengan telah diselesaikannya Skripsi ini. Penulis mempersembahkannya kepada :

1. Tuhan Yang Maha Esa ALLAH SWT, yang atas rahmat dan rizki-Nya memberikan berkah ilmu dan wawasan yang tak terhingga.
2. Kedua orang tua saya, Bapak Sugito dan Ibu Suelmi yang selalu memberikan rasa sayang, dukungan moral dan bimbingan yang tak ternilai harganya.
3. Kakak tercinta saya Sri Ningsih yang selalu memberikan dorongan dan motivasi untuk segera menyelesaikan skripsi ini.
4. Seluruh mahasiswa Program Studi Informatika Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
5. Keluarga besar Uwuhnisty yang selalu memberikan nasihat dan arahan.
6. Keluarga besar kelas Informatika 1 angkatan 2017.
7. Sahabat saya yang tak mungkin bisa saya sebutkan satu persatu, terimakasih atas semua dukungan dan semangatnya.
8. Segenap Civitas akademik Universitas AMIKOM Yogyakarta.
9. Dan untuk semua pihak yang telah bertanya “Kapan Sidang?”, “Kapan Wisuda?” dan lain sebagainya.

KATA PENGANTAR

Dengan memanjatkan puja dan puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat, taufik dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan judul **“Sistem Klasifikasi Faktor Resiko Diabetes Mellitus Menggunakan Metode Decision Tree C4.5”**, sebagai salah satu syarat dalam mencapai Gelar Sarjana di Program Studi Informatika Universitas AMIKOM Yogyakarta.

Penulis menyadari bahwa skripsi ini tidak mungkin terselesaikan tanpa adanya dukungan, bantuan, bimbingan dan nasehat dari berbagai pihak selama penyusunan skripsi ini. Pada kesempatan ini penulis menyampaikan terimakasih setulus-tulusnya kepada :

1. Bapak Prof. Dr. M. Suyanto, MM., selaku rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradnya Dhunita, M.Kom., selaku Ketua Program Studi Informatika Universitas AMIKOM Yogyakarta.
4. Ibu Erni Seniwati S.Kom, M.Cs., selaku dosen pembimbing yang dengan sabar memberikan arahan dan bimbingan dari awal hingga akhir proses penyusunan skripsi ini.
5. Kedua orang tua penulis, Bapak Sugito dan Ibu Suelmi serta kakak kandung saya Sri Ningsih yang selalu memberikan dukungan, doa, kasih sayang serta motivasi dalam penyusunan skripsi ini.
6. Keluarga besar Danis Salon yang telah memberikan semangat dalam pengerjaan skripsi ini.

Demikian ucapan terima kasih yang dapat penulis sampaikan kepada seluruh pihak, semoga Tuhan Yang Maha Esa membalas kebaikan mereka dan semoga skripsi ini dapat bermanfaat bagi para penulis dan seluruh pihak yang membaca. Aamiin...

Yogyakarta, 29 Maret 2023




Agus Triyanto

DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN	v
MOTTO.....	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiv
INTISARI.....	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Metodologi Penelitian	3
1.5.1 Metode Pengumpulan Data.....	3
1.5.2 Tahap Tahap Penelitian	4
1.6 Sistematika Penulisan.....	5
BAB II LANDASAN TEORI	6
2.1 Tinjauan Pustaka.....	6
2.2 Pengertian Diabetes Melitus.....	8

2.3	Data Mining.....	8
2.4	Decision Tree.....	9
2.5	Algoritma C4.5.....	10
2.6	Kaggle.....	12
2.7	Validasi dan Evaluasi.....	12
BAB III METODE PENELITIAN.....		15
3.1	Tahapan Persiapan.....	15
3.2	Analisa Kebutuhan Penelitian.....	15
3.2.1	Analisa Kebutuhan Fungsional.....	15
3.2.2	Analisa Kebutuhan Non-Fungsional.....	15
3.3	Dataset.....	16
3.4	Pengolahan Data.....	16
3.5	Deskripsi Sistem.....	17
3.6	Kebutuhan Data.....	19
3.7	Representasi Data Menggunakan Metode Decision Tree C4.5.....	44
3.7.1	Perhitungan Pada Data Latih.....	44
3.8	Implementasi Sistem.....	83
3.9	Pengujian Sistem.....	83
3.10	Evaluasi Sistem.....	83
3.11	Kesimpulan dan Saran.....	83
BAB IV HASIL DAN PEMBAHASAN.....		85
4.1	Skenario Pengujian.....	85
4.2	Evaluasi Sistem.....	86
4.3	Hasil Pengujian.....	87
4.3.1	Pembahasan hasil pengujian pertama hingga pengujian kelima.....	90



4.4	Hasil Pengujian Akurasi	91
4.5	White Box Testing	92
4.6	Black Box Testing	93
4.7	Implementasi Sistem	93
4.7.1	Halaman Login.....	93
4.7.2	Halaman Home.....	94
4.7.3	Halaman Olah Data	94
4.7.4	Halaman Data Mining.....	95
4.7.5	Halaman Pohon Keputusan	95
4.7.6	Halaman Uji Rule.....	96
4.7.7	Halaman User	96
BAB V PENUTUP.....		98
5.1	Kesimpulan.....	98
5.2	Saran	98
DAFTAR PUSTAKA		99
LAMPIRAN.....		101

DAFTAR TABEL

Tabel 2. 1 Perbandingan Tinjauan Pustaka	7
Tabel 2. 2 Confusion Matrix	13
Tabel 3. 1 Data yang diperoleh dari data record	19
Tabel 3. 2 Data latih setelah dilakukan preprocessing	22
Tabel 3. 3 Data uji setelah dilakukan preprocessing	37
Tabel 3. 4 Hasil perhitungan nilai gain seluruh atribut	46
Tabel 3. 5 Perhitungan percabangan node 1.1 Riwayat Darah Tinggi Ya	47
Tabel 3. 6 Perhitungan percabangan node 1.2 Riwayat Darah Tinggi Tidak	48
Tabel 3. 7 Perhitungan nilai gain pada percabangan node 1.1.1 “Riwayat Darah Tinggi Ya, IMT Kurang”	50
Tabel 3. 8 Perhitungan nilai gain pada percabangan node 1.1.2 “Riwayat Darah Tinggi Ya, IMT Normal”	51
Tabel 3. 9 Perhitungan nilai gain pada percabangan node 1.1.3 “Riwayat Darah Tinggi Ya, IMT Berlebih”	53
Tabel 3. 10 Perhitungan nilai gain pada percabangan node 1.1.3.1 “Riwayat Darah Tinggi Ya, IMT Berlebih, Riwayat Kolestrol Tinggi Ya”	54
Tabel 3. 11 Perhitungan nilai gain pada percabangan node 1.1.4 “Riwayat Darah Tinggi Ya, IMT Obesitas”	55
Tabel 3. 12 Perhitungan nilai gain pada percabangan node 1.1.4.1 “Riwayat Darah Tinggi Ya, IMT Obesitas, Kelamin Laki-laki”	56
Tabel 3. 13 Perhitungan nilai gain pada percabangan node 1.1.4.2 “Riwayat Darah Tinggi Ya, IMT Obesitas, Kelamin Perempuan”	57
Tabel 3. 14 Perhitungan nilai gain pada percabangan node 1.1.4.1.3 “Riwayat Darah Tinggi Ya, IMT Obesitas, Kelamin Laki-laki, Rentang Usia Lansia”	58
Tabel 3. 15 Perhitungan nilai gain pada percabangan node 1.1.4.1.4 “Riwayat Darah Tinggi Ya, IMT Obesitas, Kelamin Laki-laki, Rentang Usia Manula”	59
Tabel 3. 16 Perhitungan nilai gain pada percabangan node 1.1.4.1.4.2 “Riwayat Darah Tinggi Ya, IMT Obesitas, Kelamin Laki-laki, Rentang Usia Manula, Peminum Alkohol Berat Tidak”	60
Tabel 3. 17 Perhitungan nilai gain pada node 1.1.4.1.4.2.1 “Riwayat Darah Tinggi Ya, IMT Obesitas, Kelamin Laki-laki, Rentang Usia Manula, Peminum Alkohol Berat Tidak, Riwayat Kolestrol Tinggi Ya”	61
Tabel 3. 18 Perhitungan nilai gain pada node 1.1.4.1.4.2.2 “Riwayat Darah Tinggi Ya, IMT Obesitas, Kelamin Laki-laki, Rentang Usia Manula, Peminum Alkohol Berat Tidak, Riwayat Kolestrol Tinggi Tidak”	61
Tabel 3. 19 Perhitungan nilai gain pada node 1.1.4.1.4.2.1.1 “Riwayat Darah Tinggi Ya, IMT Obesitas, Kelamin Laki-laki, Rentang Usia Manula, Peminum Alkohol Berat Tidak, Riwayat Kolestrol Tinggi Ya, Riwayat Jantung Koroner”	62

Tabel 3. 20 Perhitungan nilai gain pada node 1.1.4.2.2 “Riwayat Darah Tinggi Ya, IMT Obesitas, Kelamin Perempuan, Peminum Alkohol Berat Tidak”	63
Tabel 3. 21 Perhitungan nilai gain pada node 1.1.4.2.2 “Riwayat Darah Tinggi Ya, IMT Obesitas, Kelamin Perempuan, Peminum Alkohol Berat Tidak, Rentang Usia Dewasa”	64
Tabel 3. 22 Perhitungan nilai gain pada node 1.2.2 “Riwayat Darah Tinggi Tidak, Rentang Usia Dewasa”	65
Tabel 3. 23 Perhitungan nilai gain pada node 1.2.2 “Riwayat Darah Tinggi Tidak, Rentang Usia Lansia”	67
Tabel 3. 24 Perhitungan nilai gain pada node 1.2.3.4 “Riwayat Darah Tinggi Tidak, Rentang Usia Lansia, IMT Obesitas”	68
Tabel 3. 25 Perhitungan nilai gain pada node 1.2.3.4.2 “Riwayat Darah Tinggi Tidak, Rentang Usia Lansia, IMT Obesitas, Peminum Alkohol Berat Tidak”	69
Tabel 3. 26 Perhitungan nilai gain pada node 1.2.3.4.2.2 “Riwayat Darah Tinggi Tidak, Rentang Usia Lansia, IMT Obesitas, Peminum Alkohol Berat Tidak, Riwayat Merokok Tidak”	69
Tabel 3. 27 Perhitungan nilai gain pada node 1.2.3.4.2.2.1 “Riwayat Darah Tinggi Tidak, Rentang Usia Lansia, IMT Obesitas, Peminum Alkohol Berat Tidak, Riwayat Merokok Tidak, Kelamin Laki-laki”	70
Tabel 3. 28 Perhitungan nilai gain pada node 1.2.3.4.2.2.2 “Riwayat Darah Tinggi Tidak, Rentang Usia Lansia, IMT Obesitas, Peminum Alkohol Berat Tidak, Riwayat Merokok Tidak, Kelamin Perempuan”	70
Tabel 3. 29 Perhitungan nilai gain pada percabangan node 1.2.4 “Riwayat Darah Tinggi Tidak, Rentang Manula”	72
Tabel 3. 30 Perhitungan nilai gain pada percabangan node 1.2.4.1 “Riwayat Darah Tinggi Tidak, Rentang Manula, Peminum Alkohol Berat Ya”	73
Tabel 3. 31 Perhitungan nilai gain pada percabangan node 1.2.4.2 “Riwayat Darah Tinggi Tidak, Rentang Manula, Peminum Alkohol Berat Tidak”	73
Tabel 3. 32 Hasil prediksi pada data uji	81
Tabel 4. 1 Perhitungan Confusion Matrik pada Algoritma Decision Tree C4.5...	86
Tabel 4. 2 Hasil Pengujian Pertama	87
Tabel 4. 3 Hasil Pengujian Kedua	88
Tabel 4. 4 Hasil Pengujian Ketiga	88
Tabel 4. 5 Hasil Pengujian Keempat	89
Tabel 4. 6 Hasil Pengujian Kelima	90
Tabel 4. 7 Black Box Testing	93

DAFTAR GAMBAR

Gambar 3. 1 Flowchart Sistem Klasifikasi Faktor Resiko Diabetes Mellitus menggunakan metode Decision Tree C4.5	18
Gambar 3. 2 Hasil Pembentukan cabang pada node seluruh data.....	47
Gambar 3. 3 Hasil pembentukan cabang pada node 1.1 “Riwayat Darah Tinggi Ya” dan node 1.2 “Riwayat Darah Tinggi Tidak”	49
Gambar 3. 4 Hasil pembentukan cabang node 1.1.1 “Riwayat Darah Tinggi Ya, IMT Kurus”	51
Gambar 3. 5 Hasil pembentukan cabang pada node 1.1.2 “Riwayat Darah Tinggi Ya, IMT Normal”	52
Gambar 3. 6 Hasil pembentukan cabang pada node 1.1.3	55
Gambar 3. 7 Hasil pembentukan cabang pada keseluruhan node 1.1.4.....	65
Gambar 3. 8 Hasil pembentukan cabang node 1.2.2 “Riwayat Darah Tidak, Rentang Usia Dewasa”	66
Gambar 3. 9 Hasil pembentukan cabang pada node 1.2.3	71
Gambar 3. 10 Pohon Keputusan yang Terbentuk	75
Gambar 4. 1 Hasil Pengujian Akurasi.....	92
Gambar 4. 2 Halaman Login.....	94
Gambar 4. 3 Halaman Home.....	94
Gambar 4. 4 Halaman Olah Data	95
Gambar 4. 5 Halaman Data Mining	95
Gambar 4. 6 Halaman Pohon Keputusan	96
Gambar 4. 7 Halaman Uji Rule.....	96
Gambar 4. 8 Halaman User.....	97
Gambar 4. 9 Halaman Hasil Prediksi User	97

INTISARI

Diabetes Mellitus (DM) merupakan suatu penyakit kronis yang dapat menyebabkan penderitanya kehilangan kemampuan dalam mengelola kadar gula darah didalam tubuh dan dapat mempengaruhi kualitas dan harapan hidup. Secara umum terdapat 2 tipe Diabetes Mellitus yakni Diabetes Mellitus tipe I dan Diabetes Mellitus tipe II. Pemeriksaan dini terhadap DM diharapkan dapat mengurangi dan mencegah kenaikan DM yang akan terjadi. Karena dengan adanya pemeriksaan secara dini terhadap resiko DM tersebut, diharapkan dapat meningkatkan kesadaran masyarakat dalam mengubah pola hidup mereka agar memperkecil kemungkinan terkena resiko DM.

Penelitian ini menerapkan teknik Data Mining klasifikasi dengan menggunakan metode Decision Tree C4.5, yang merupakan salah satu metode yang dapat digunakan untuk melakukan klasifikasi faktor resiko Diabetes Mellitus dengan menghasilkan rule dalam bentuk pohon keputusan untuk menentukan kelas Diabetes pada klasifikasi faktor resiko Diabetes yaitu kelas Diabetes dan Non Diabetes. Atribut yang digunakan terdiri dari 10 atribut yakni Riwayat Darah Tinggi, Riwayat Kolestrol Tinggi, Index Massa Tubuh (IMT), Riwayat Merokok, Riwayat Stroke, Riwayat Jantung Koroner, Aktifitas Fisik, Riwayat Peminum Alkohol Berat, Jenis Kelamin dan Rentang Usia.

Pengujian sistem ini dilakukan 5 kali pengujian. Dari hasil pengujian didapatkan akurasi tertinggi algoritma Decision Tree C4.5 mendapatkan akurasi terbaik pada pengujian pertama dengan akurasi mencapai 96,33%. Total jumlah data yang digunakan pada pengujian pertama yaitu 300 record data dengan pembagian 70% sebagai data latih dan 30% sebagai data uji. Algoritma Decision Tree C4.5 dapat memprediksi resiko Diabetes Mellitus berdasarkan faktor – faktor resiko pengguna. Hasil prediksi yang diperoleh yaitu kelas Diabetes dan kelas Non Diabetes. Dari data latih yang digunakan, pasien yang mempunyai riwayat stroke lebih rentan terkena faktor resiko Diabetes Mellitus dengan persentase 89.36%. Dari data latih yang digunakan, pasien yang tidak mempunyai riwayat stroke memiliki kemungkinan 56.1% tidak memiliki resiko Diabetes Mellitus.

Kata Kunci : Data Mining, Klasifikasi, Decision Tree, Diabetes Mellitus

ABSTRACT

Diabetes Mellitus (DM) is a chronic disease that can cause sufferers to lose their ability to manage blood sugar levels in the body and can affect quality and life expectancy. In general, there are 2 types of Diabetes Mellitus, namely Diabetes Mellitus type I and Diabetes Mellitus type II. Early examination of DM is expected to reduce and prevent an increase in DM that will occur. Due to the existence of an early examination of the risk of DM, it is hoped that it can increase public awareness in changing their lifestyle so as to minimize the possibility of being exposed to the risk of DM.

This study applies classification data mining techniques using the Decision Tree C4.5 method, which is a method that can be used to classify risk factors for Diabetes Mellitus by generating rules in the form of a decision tree to determine the class of Diabetes in the classification of Diabetes risk factors, namely the Diabetes class. and Non Diabetic. The attributes used consisted of 10 attributes, namely history of high blood pressure, history of high cholesterol, body mass index (BMI), history of smoking, history of stroke, history of coronary heart disease, physical activity, history of heavy alcohol consumption, gender and age range.

Testing this system is done 5 times the test. From the test results, the highest accuracy of the Decision Tree C4.5 algorithm obtained the best accuracy in the first test with an accuracy of 96.33%. The total amount of data used in the first test is 300 data records with a division of 70% as training data and 30% as test data. The Decision Tree Algorithm C4.5 can predict the risk of Diabetes Mellitus based on the user's risk factors. The prediction results obtained are the Diabetes class and the Non Diabetes class. From the training data used, patients who have a history of stroke are more susceptible to risk factors for Diabetes Mellitus with a percentage of 89.36%. From the training data used, patients who do not have a history of stroke have a 56.1% chance of not having the risk of Diabetes Mellitus.

Keywords: *Data Mining, Classification, Decision Tree, Diabetes Mellitus*