

ALGORITMA *ITERATIVE DICHOTOMISER 3* (ID3) UNTUK

MENGIDENTIFIKASI DATA REKAM MEDIS

(Studi Kasus Penyakit Diabetes Mellitus di Balai Kesehatan Kementerian

Perindustrian, Jakarta)



SKRIPSI

Oleh:

AVIA ENGGARTYASTI

NIM. 24010210141032

JURUSAN STATISTIKA

FAKULTAS SAINS DAN MATEMATIKA

UNIVERSITAS DIPONEGORO

SEMARANG

2015

**ALGORITMA *ITERATIVE DICHOTOMISER 3* (ID3) UNTUK
MENGIDENTIFIKASI DATA REKAM MEDIS
(Studi Kasus Penyakit Diabetes Mellitus di Balai Kesehatan Kementerian
Perindustrian, Jakarta)**

Oleh :

AVIA ENGGARTYASTI

NIM. 24010210141032

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Sains pada Jurusan Statistika**

**JURUSAN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG**

2015

HALAMAN PENGESAHAN I

Judul : Algoritma *Iterative Dichotomiser 3* (ID3) untuk
Mengidentifikasi Data Rekam Medis
(Studi Kasus Penyakit Diabetes Mellitus di Balai
Kesehatan Kementerian Perindustrian, Jakarta)

Nama Mahasiswa : Avia Enggar Tyasti

NIM : 24010210141032

Jurusan : Statistika

Telah diujikan pada sidang Tugas Akhir pada tanggal 13 Februari 2015 dan
dinyatakan lulus pada tanggal 25 Februari 2015.

Semarang, Februari 2015

Mengetahui,

Ketua Jurusan Statistika

Bakultas Sains dan Matematika



[Signature]
Dra. Lwi Isprivanti, M.Si
NIP. 195709141986032001

Panitia Penguji Ujian Tugas Akhir

Ketua,

Moch. Abdul Mukid, S.Si, M.Si
NIP. 197808172005011001

HALAMAN PENGESAHAN II

Judul : Algoritma *Iterative Dichotomiser 3* (ID3) untuk
Mengidentifikasi Data Rekam Medis
(Studi Kasus Penyakit Diabetes Mellitus di Balai
Kesehatan Kementerian Perindustrian, Jakarta)

Nama Mahasiswa : Avia Enggar Tyasti


NIM : 24010210141032

Jurusan : Statistika

Telah diujikan pada sidang Tugas Akhir pada tanggal 13 Februari 2015

Semarang, 26 Februari 2015

Pembimbing I



Dra. Dwi Ispriyanti, M.Si
NIP. 195709141986032001

Pembimbing II



Abdul Hoyvi, S.Si, M.Si
NIP. 197202022008011018

KATA PENGANTAR

Puji syukur penulis ucapkan atas kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “**Algoritma Iterative Dichotomiser 3 (ID3) untuk Mengidentifikasi Data Rekam Medis (Studi Kasus Penyakit Diabetes Mellitus di Balai Kesehatan Kementerian Perindustrian, Jakarta)**”. Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Ibu Dra. Dwi Ispriyanti, M.Si selaku Ketua Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro Semarang sekaligus sebagai Dosen Pembimbing I
2. Bapak Abdul Hoyyi, S.Si., M.Si. sebagai pembimbing II yang telah memberikan bimbingan dan pengarahan dalam penulisan Tugas Akhir ini
3. Bapak dan Ibu dosen Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro yang telah memberikan ilmu yang bermanfaat.
4. Semua pihak yang tidak dapat disebutkan satu per satu yang telah mendukung penulis menyelesaikan penulisan Tugas Akhir ini.

Penulis berharap Tugas Akhir ini dapat bermanfaat bagi seluruh civitas akademika di Universitas Diponegoro, khususnya Jurusan Statistika dan masyarakat pada umumnya.

Semarang, Februari 2015

Penulis

ABSTRAK

Algoritma *Iterative Dichotomiser 3* (ID3) adalah algoritma *decision tree learning* (algoritma pembelajaran pohon keputusan) yang paling dasar. Algoritma ini melakukan pencarian secara menyeluruh (*greedy*) pada semua kemungkinan pohon keputusan. Algoritma ID3 dapat diimplementasikan menggunakan fungsi *rekursif* (fungsi yang memanggil dirinya sendiri). Salah satu permasalahan yang dapat diselesaikan menggunakan algoritma ID3 adalah klasifikasi pasien diabetes. Diabetes adalah suatu penyakit karena tubuh tidak mampu mengendalikan jumlah gula atau glukosa dalam aliran darah. Klasifikasi menggunakan ID3 pada kasus penderita diabetes menghasilkan pohon dengan banyak simpul mencapai 32 simpul dimana 21 diantaranya adalah simpul daun dan atribut glukosa puasa dua jam *postprandial* terpilih sebagai simpul akar dalam pembuatan pohon keputusan. Berdasarkan pengukuran kinerja klasifikasi menunjukkan bahwa akurasi atau ukuran ketepatan klasifikasi mencapai 89,75%. Sedangkan pengukuran akurasi hasil klasifikasi Algoritma ID3 menggunakan sampel pengujian yang berjumlah 84 sampel menunjukkan akurasi sebesar 72,619%.

Kata Kunci : Algoritma ID3, Pohon Keputusan, Diabetes

ABSTRACT

Iterative Dichotomiser 3 (ID3) Algorithm is a basic decision tree learning algorithm. These algorithms perform a thorough search (greedy) in all possible decision tree. ID3 algorithm can be implemented using a recursive function, (function that calls itself). One of the problems that can be solved using the ID3 algorithm is a classification of diabetic patients. Diabetic is a disease because of the body is not able to control the amount of sugar or glucose in the bloodstream. Classification using ID3 in the case of diabetics produce trees with many vertices to 32 knot where 21 of them is a leaf node and attribute two-hour postprandial glucose fasting elected as the root node in the decision-making tree. Based on the classification performance measurements show that the classification accuracy or measurement accuracy reaches 89,75%. While the measurement accuracy of the classification algorithm ID3 using test samples totaling 84 samples showed an accuracy of 72,619%

Keywords: ID3 Algorithm, Decision Tree, Diabetes

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN I	ii
HALAMAN PENGESAHAN II.....	iii
KATA PENGANTAR	iv
ABSTRAK.....	v
ABSTRACT.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN.....	xi
BAB I PENDAHULUAN	
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	4
1.4. Tujuan	5
BAB II TINJAUAN PUSTAKA	
2.1. Data Mining	6
2.2. <i>Decision Tree</i>	7
2.3. <i>Algoritma Iterative Dichotomiser 3 (ID3)</i>	9
2.3.1 <i>Entropy</i>	11
2.3.2 <i>Information Gain</i>	12

2.3.3	Ketepatan Pohon Klasifikasi	12
2.4.	Diabetes Mellitus	13
2.4.1	Penyebab Diabetes	14
2.4.2	Alat Ukur	15
BAB III METODOLOGI PENELITIAN		
3.1.	Jenis dan Sumber Data	16
3.2.	Variabel Penelitian	16
3.3.	Langkah-langkah Analisis	16
3.4.	Tahapan Alur penelitian	18
BAB IV HASIL DAN PEMBAHASAN		
4.1.	Statistika Deskriptif	21
4.2.	Algoritma <i>Iterative Dichotomiser 3</i> (ID3)	29
4.2.1	Konstruksi Algoritma ID3	29
4.2.2	Analisis Pohon Keputusan.....	33
4.2.3	Pengukuran Ketepatan Hasil Pohon Klasifikasi.....	34
4.2.4	Identifikasi Status Diabetes Pasien	35
4.2.5	Hasil Pohon Keputusan Dengan Data Testing	38
BAB V KESIMPULAN		39
DAFTAR PUSTAKA		40
LAMPIRAN		42

DAFTAR TABEL

	Halaman
Tabel 1. Matriks Konfusi	13
Tabel 2. Kriteria Jenis Kelamin dan Usia Pasien	19
Tabel 3. Kriteria Diabetes Mellitus	20
Tabel 4. Status Diabetes Pasien.....	22
Tabel 5. Status Diabetes Pasien Berdasarkan Jenis Kelamin.....	23
Tabel 6. Status Diabetes Pasien Berdasarkan Usia	24
Tabel 7. Status Diabetes Pasien Berdasarkan Glukosa Puasa.....	25
Tabel 8. Status Diabetes Pasien Berdasarkan Glukosa 2 Jam PP	26
Tabel 9. Status Diabetes Pasien Berdasarkan Kadar Trygliserida	26
Tabel 10. Status Diabetes Pasien Berdasarkan Kadar HDL.....	27
Tabel 11. Status Diabetes Pasien Berdasarkan LDL.....	28
Tabel 12. Status Diabetes Pasien Berdasarkan kadar hbA1c	28
Tabel 13. Frekuensi Tiap Kelas (Positif dan Negatif).....	29
Tabel 14. Proporsi Masing-masing Kelas	29
Tabel 15. Frekuensi Masing-masing Kategori pada Atribut Glukosa 2 Jam PP Berdasarkan Kelasnya	30
Tabel 16. Proporsi Masing-masing Kategori Berdasarkan Kelasnya	30
Tabel 17. Nilai <i>Informartion Gain</i>	31
Tabel 18. Hasil Matriks Konfusi Algoritma ID3 dengan Data <i>Training</i>	33
Tabel 19. Hasil Matriks Konfusi Sampel Pengujian dengan Data <i>Testing</i> ...	37

DAFTAR GAMBAR

	Halaman
Gambar 1. Diagram Alur Penelitian	18
Gambar 2. Status Diabetes pasien.....	22
Gambar 3. Status Diabetes Pasien Berdasarkan Jenis Kelamin.....	23
Gambar 4. Pohon Keputusan Tingkat Pertama.....	32

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Data Rekam Medis Pasien Diabetes.....	41
Lampiran 2. Hasil Algoritma ID3 Menggunakan Data <i>Training</i>	42
Lampiran 3. Pohon Keputusan yang Terbentuk Dengan Data <i>Training</i>	46
Lampiran 4. Hasil Algoritma ID3 Menggunakan Data <i>Testing</i>	47
Lampiran 5. Hasil Pengklasifikasian Menggunakan Data <i>Testing</i>	49

BAB I

PENDAHULUAN

1.1 Latar Belakang

Algoritma *Iterative Dichotomiser 3* (ID3) merupakan salah metode dalam data mining. Data Mining mulai dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang akademi, bisnis hingga medis. ID3 adalah algoritma *decision tree learning* (algoritma pembelajaran pohon) yang paling dasar. Algoritma ini melakukan pencarian secara menyeluruh pada semua kemungkinan pohon keputusan. Algoritma ID3 dapat diimplementasikan menggunakan fungsi rekursif (fungsi yang dapat memanggil dirinya sendiri). Pembentukan pohon klasifikasi dengan algoritma ID3 melalui dua langkah, yaitu menghitung nilai *entropy* dan menghitung nilai *information gain* dari setiap variabel. ID3 dapat menyelesaikan kasus pada berbagai bidang salah satunya dapat diterapkan pada bidang kesehatan (Santosa, 2007).

Kesehatan merupakan aspek penting dalam kehidupan, banyak permasalahan yang terjadi dalam peningkatan taraf kesehatan masyarakat sehubungan gaya hidup yang kurang sehat (*unhealthy lifestyle*), akibat buruk dari *unhealthy lifestyle* dapat berujung pada munculnya berbagai macam penyakit. Penyakit metabolik adalah salah satu contoh dampak buruk dari *unhealthy lifestyle*. Masalah yang sering terjadi dalam gaya hidup masyarakat tersebut adalah Diabetes Mellitus (DM) yang merupakan penyakit yang disebabkan kadar

gula darah yang tinggi. Hal ini menjadi tantangan yang berat pada sistem pelayanan kesehatan di negeri ini (Zahtamal, 2007).

DM pada saat ini merupakan salah satu masalah kesehatan yang berdampak pada produktivitas dan menurunkan mutu sumber daya manusia. Berdasarkan catatan organisasi kesehatan dunia tahun 1998, Indonesia menduduki peringkat keenam dengan jumlah penderita diabetes terbanyak setelah India, Cina, Rusia, Jepang dan Brazil. Penderita DM di Indonesia semakin meningkat, hal ini dapat diketahui bahwa pada tahun 1995 terdapat lebih kurang 5 juta penderita DM di Indonesia dengan peningkatan sekitar 230 ribu penderita setiap tahun sehingga pada tahun 2025 penderita *Diabetes* di Indonesia diperkirakan akan mencapai 12 juta orang. Peningkatan terjadi akibat bertambahnya populasi penduduk usia lanjut dan perubahan gaya hidup mulai dari pola makan/jenis makanan yang dikonsumsi sampai berkurangnya kegiatan jasmani. Hal ini terjadi terutama pada kelompok usia dewasa ke atas pada seluruh status sosial ekonomi. Selain itu peningkatan jumlah kasus DM terjadi karena kurangnya tenaga kesehatan, peralatan pemantauan dan obat-obatan tertentu, terutama di daerah terpencil serta belum ada keseragaman dalam mengelola pasien DM oleh dokter di lini depan (Zahtamal, 2007).

Banyak penyandang penyakit diabetes yang terlambat memperoleh penanganan yang diakibatkan terlambatnya identifikasi pada pasien tersebut, padahal apabila dilakukan diagnosis secara dini, maka penanganan bisa dilakukan lebih cepat dan hal yang membahayakan dapat dihindari. Selain jenis kelamin dan usia klasifikasi seseorang dinyatakan mengidap penyakit DM adalah melalui beberapa tes kesehatan seperti pengecekan glukosa darah puasa, glukosa darah

dua jam sesudah makan, kadar HDL (*High Density Lipoprotein*), kadar LDL (*Low Density Lipoprotein*), *triglyserida* dan juga melalui tes hbA1c, oleh karena itu dibutuhkan sistem yang dapat mengidentifikasi penyakit diabetes agar penyakit dapat diketahui secara cepat, tepat dan sedini mungkin.

Beberapa metode yang sering digunakan dalam pengklasifikasian adalah Analisis Diskriminan, Regresi Logistik Biner, algoritma *Iterative Dichotomiser 3* (ID3) dan lain-lain. Untuk mengidentifikasi penyakit Diabetes Mellitus tersebut, perlu diketahui ciri-ciri pasien penyakit Diabetes Mellitus melalui berbagai hasil pengecekan tes laboratorium. Hasil pengecekan tersebut memiliki nilai diskret yang dapat dikategorikan, sehingga pada penelitian ini metode statistik klasifikasi yang digunakan adalah algoritma *Iterative Dichotomiser 3* (ID3).

Beberapa metode statistika yang telah digunakan pada penelitian sebelumnya pada kasus DM dan algoritma *Iterative Dichotomiser 3* (ID3) antara lain adalah “*Faktor-faktor Mempengaruhi Terjadinya Ulkus Diabetikum Pada Pasien Diabetes Melitus Tipe 2 Di RSUD Prof. DR. Margono Soekarjo Purwokerto*” Oleh Ferawati (2014), “*Hubungan Antara Diabetes Melitus Tipe 2 Dengan Retinopati Diabetik Dikaji Dari HbA1c Sebagai Parameter Kontrol Gula Darah*” oleh Rangkuti (2011) dan “*Klasifikasi Jurnal Ilmiah Berbahasa Inggris Berdasarkan Abstrak Menggunakan Algoritma ID3*” oleh Wijakso (2013).

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang penelitian, permasalahan yang diangkat pada penelitian ini adalah

1. Bagaimana penerapan algoritma *Iterative Dichotomiser 3* (ID3) dalam pembentukan pohon klasifikasi untuk mengetahui ciri-ciri pasien penyakit Diabetes Mellitus dengan menggunakan hasil tes laboratorium.
2. Bagaimana hasil klasifikasi yang dihasilkan dari pohon klasifikasi menggunakan algoritma *Iterative Dichotomiser 3* (ID3), sehingga dapat diketahui hasil akurasi pohon klasifikasi dalam mengklasifikasi.

1.3 Batasan Masalah

Pada penelitian ini, dilakukan pembatasan masalah yaitu

1. Penelitian ini menggunakan metode algoritma *Iterative Dichotomiser 3* (ID3).
2. Hasil klasifikasi dibagi menjadi dua, yaitu positif diabetes dan negatif diabetes.
3. Dalam mengklasifikasi untuk menentukan ciri-ciri pasien penyakit Diabetes Mellitus digunakan variabel jenis kelamin pasien, umur pasien, glukosa darah puasa, glukosa darah dua jam sesudah makan, kadar HDL (*High Density Lipoprotein*), kadar LDL (*Low Density Lipoprotein*), kadar *triglyserida* dan kadar hbA1c.

1.4 Tujuan

Tujuan penulisan tugas akhir ini adalah

1. Membentuk pohon klasifikasi untuk mengetahui ciri-ciri pasien penyakit Diabetes Mellitus dengan hasil tes laboratorium menggunakan metode algoritma *Iterative Dichotomiser 3* (ID3).
2. Mengetahui tingkat akurasi pohon klasifikasi dalam mengklasifikasi data.