

Penentuan Menu Makanan Untuk Penderita Diabetes Menggunakan Metode *Iterative Dichotomizer Tree* (ID3)

Naufal Sakagraha Kuspinta¹, Agus Wahyu Widodo², M. Tanzil Furqon³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹nsakagraha@gmail.com, ²a_wahyu_w@ub.ac.id, ³m.tanzil.furqon@ub.ac.id

Abstrak

Diabetes Melitus (DM) merupakan kelainan metabolik yang disebabkan oleh beberapa faktor seperti kurangnya insulin atau ketidakmampuan tubuh untuk memanfaatkan insulin. Sebagian besar penderita mengidap penyakit ini akibat faktor keturunan dan pola hidup yang tidak sehat. Diabetes Melitus juga merupakan penyakit menahun yang menjadi masalah kesehatan masyarakat di Indonesia. Pada tubuh penderita DM, ketidakmampuan mengatur gula darah secara otomatis seperti manusia normal menjadikannya menderita penyakit diabetes. Peningkatan yang signifikan atas hiperglikemia atau kadar gula darah tersebut diyakini bertambah seiring dengan konsumsi yang dicerna. Penyebab dasarnya tentu dari sumber makanan yang dikonsumsi. *Iterative Dichotomizer Tree* (ID3) merupakan salah satu metode klasifikasi. Subjek yang terdapat pada penelitian ini adalah aplikasi pengelompokan data dengan menggunakan *Iterative Dichotomizer Tree* (ID3) untuk mengelompokkan data menu makanan penderita diabetes. *Dataset* yang digunakan pada penelitian ini bersumber dari Puskesmas Kendalsari Malang. Hasil dari penelitian ini adalah sebuah sistem yang mampu melakukan pengelompokan *dataset* menu makanan bagi penderita diabetes dengan menggunakan metode ID3. Dari pengujian yang dilakukan didapatkan hasil akurasi dengan kombinasi data latih 80 dan data uji 20 sebesar 75%.

Kata kunci: *Diabetes Mellitus, Iterative Dichotomizer Tree* (ID3), Menu Makanan

Abstract

Diabetes Mellitus (DM) is a metabolic disorder caused by several factors such as insulin deficiency or the inability of the body to utilize insulin. Most patients suffer from this disease due to heredity and unhealthy lifestyle. Diabetes Mellitus is also a chronic disease that became a public health problem in Indonesia. In the body of patients with DM, the inability to automatically adjust the sugar levels in the blood like a healthy person makes them suffering from diabetes. Significant increases in hyperglycemia or sugar levels are believed to increase along with digested consumption. Causes of food sources consumed. Iterative Dichotomizer Tree (ID3) is one of the methods. Subjects in this research is the application of data grouping by using Iterative Dichotomizer Tree (ID3) to classify diabetic food menu data. The dataset used in this study is sourced from Puskesmas Kendalsari Malang. The results of this study is a system capable of grouping food menu datasets for diabetics using the ID3 method. From the test conducted by the results of the testing of 80 data training dan 20 data testing is 75%.

Keywords: *Diabetes Mellitus, Data Mining, Iterative Dichotomizer Tree* (ID3), Food Menu

1. PENDAHULUAN

Diabetes Mellitus (DM) merupakan kelainan metabolik yang disebabkan oleh beberapa faktor seperti kurangnya insulin atau ketidakmampuan tubuh untuk memanfaatkan insulin. Sebagian besar penderita mengidap penyakit ini akibat faktor keturunan dan pola hidup yang tidak sehat. Diabetes Melitus juga

merupakan penyakit menahun yang menjadi masalah kesehatan masyarakat di Indonesia (Nati, 2013). Menurut data dari Dinas Kesehatan Kota Malang, pada tahun 2015 terdapat 5.905 kasus baru penderita DM di kota Malang, dimana 2.472 merupakan penderita laki-laki dan 3.433 merupakan penderita perempuan.

DM dikenal sebagai penyakit kencing manis atau penyakit gula darah adalah golongan

penyakit kronis yang ditandai adanya peningkatan kadar gula dalam darah (hyperglycemia) sebagai akibat dari gangguan sistem metabolisme dalam tubuh, dimana organ pancreas tidak mampu memproduksi hormon insulin sesuai kebutuhan tubuh. Pada tubuh penderita DM, ketidakmampuan mengatur gula darah secara otomatis seperti manusia normal menjadikannya menderita penyakit diabetes. Peningkatan yang signifikan atas hiperglikemia atau kadar gula darah tersebut diyakini bertambah seiring dengan konsumsi yang dicerna. Penyebab dasarnya tentu dari sumber makanan yang dikonsumsi (Waspadji, 2007).

Dalam penyakit DM terdapat 4 atribut sebagai acuan untuk menghitung kadar gula dalam darah pasien. Atribut-atribut tersebut nantinya akan membentuk suatu pola dengan kelas menu makanan yang sudah ada. Dari data tersebut, dokter dapat mendeteksi pola dengan menggunakan beberapa metode, salah satunya menggunakan *data mining*. *Data mining* merupakan metode untuk menganalisis pola atau pengetahuan dari sekumpulan data secara otomatis. Teknik *data mining* yang dapat digunakan untuk klasifikasi diantaranya adalah teknik clustering dan teknik klasifikasi (Sujana, 2010).

Teknik klasifikasi adalah proses untuk menemukan model atau fungsi yang sama pada sebuah himpunan objek dalam sebuah basis data dan mengelompokkan atau mengklasifikasikan ke dalam beberapa kelas model klasifikasi yang sudah ditetapkan. Beberapa teknik klasifikasi antara lain yaitu decision tree, Rules Base, K-Nearest Neighbor, Naïve Bayes, dan Simple Logistic. Metode yang digunakan pada penelitian ini yaitu decision tree. Decision tree adalah sistem prediksi menyerupai struktur pohon yang bercabang atau struktur hirarki, sehingga sistem ini cocok untuk menggambarkan suatu persoalan dan mencari solusi dari persoalan tersebut (Wahyudi, 2009).

2. DATA PENELITIAN

Data yang digunakan dalam penelitian ini adalah *dataset* pasien penderita diabetes yang berada di Puskesmas Kendalsari Malang. Parameter penentuan menu makanan untuk penderita diabetes yang digunakan berjumlah 4 parameter yaitu tinggi badan, berat badan, umur, dan aktifitas.

Data pasien penderita diabetes yang

diperoleh untuk penelitian ini berjumlah total 100 data. Data yang digunakan merupakan data kejadian pasien pada tahun 2015 – 2016.

3. DIABETES MELLITUS

Beberapa sumber menjelaskan pengertian Diabetes Mellitus (DM) dalam tesis Kusniawati (2011) seperti yang diuraikan berikut ini:

Menurut World Health Organization, DM sebagai suatu kelainan metabolik yang disebabkan oleh berbagai etiologi dan dimanifestasikan dengan keadaan hiperglikemia kronis dan terjadi gangguan sekresi insulin, kerja insulin atau keduanya. Menurut American Diabetes Association, DM merupakan suatu kelompok penyakit metabolik dengan karakteristik hiperglikemia yang terjadi karena kelainan sekresi insulin, kerja insulin atau kedua-duanya. Sedangkan menurut Black & Hawks (2009) mendefinisikan DM merupakan suatu penyakit kronis yang ditandai oleh ketidakmampuan tubuh untuk melakukan metabolisme karbohidrat, lemak, dan protein sehingga menyebabkan hiperglikemia. Berdasarkan beberapa definisi di atas dapat disimpulkan bahwa DM merupakan suatu penyakit metabolik yang ditandai dengan peningkatan glukosa darah (hiperglikemia) yang disebabkan oleh gangguan sekresi insulin atau kerja insulin tidak kuat yang dapat menimbulkan berbagai komplikasi.

4. ITERATIVE DICHOTOMIZER TREE

Algoritma ID3 merupakan algoritma yang paling dasar dalam decision tree dan merupakan metode untuk membangun decision tree dalam mencari solusi. Algoritma ID3 melakukan pencarian secara menyeluruh pada semua kemungkinan pohon.

Algoritma ID3 dapat diimplementasikan menggunakan fungsi rekursif (fungsi yang memanggil dirinya sendiri). Algoritma ID3 berusaha membangun decision tree secara top-down. Awalnya memeriksa semua atribut yang sesuai untuk diletakkan pada root dengan mengevaluasi semua atribut yang ada dengan menggunakan ukuran statistik (yang umum digunakan adalah information gain) untuk mengukur efektivitas suatu atribut dalam mengklasifikasikan kumpulan sampel data. Pemilihan atribut dengan menggunakan information gain, yaitu:

1. Memilih atribut dengan information gain terbesar.

2. Buat Simpul yang berisi atribut tersebut.
3. Proses perhitungan information gain akan terus dilaksanakan sampai semua data telah termasuk dalam kelas yang sama. Atribut yang telah dipakai tidak disertakan lagi dalam perhitungan nilai information gain.

Dalam perhitungan algoritma ID3 dibutuhkan 2 rumus utama yaitu entropy dan information gain. Fungsi dari entropy yaitu mengukur seberapa baiknya sebuah node dan digunakan sebagai suatu parameter untuk mengukur keberagaman dari kumpulan sampel data. Ketika entropy semakin besar, menunjukkan bahwa kumpulan data semakin heterogen. Fungsi gain yaitu mengukur seberapa baik suatu atribut memisahkan training example ke dalam kelas target. Entropy digunakan untuk mendefinisikan nilai information gain yang berguna untuk memilih atribut terbaik dari atribut-atribut yang tersedia. Berikut perhitungan algoritma ID3 dengan mencari nilai dari entropy dan information gain:

$$Entropy(S) = -P_+ \log_2 P_+ - P_- \log_2 P_- \quad (1)$$

Keterangan:

- E(S) adalah entropy dari himpunan S
- P₊ adalah probabilitas sampel S yang mempunyai class positif. P₊ dihitung dengan membagi jumlah sample positif (S₊) dengan jumlah sampel keseluruhan (S) sehingga $P_+ = \frac{S_+}{S}$
- P₋ adalah probabilitas sampel S yang mempunyai class negative. P₋ dihitung dengan membagi jumlah sample negative (S₋) dengan jumlah sampel keseluruhan (S) sehingga $P_- = \frac{S_-}{S}$

Pada algoritma ID3 pengurangan entropy disebut dengan information gain. Pembagian sampel S terhadap atribut X dapat dihitung dengan rumus (2):

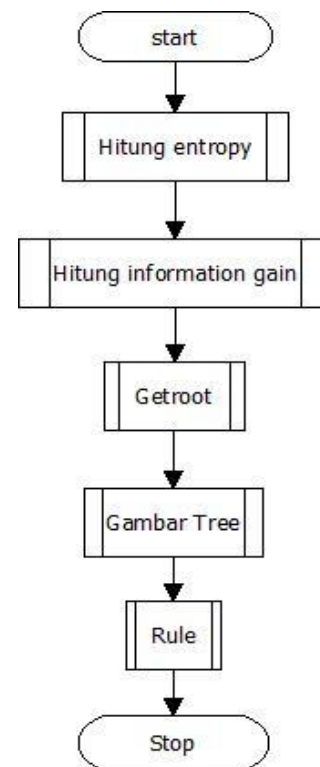
$$Gain(S, X) = Entropy(S) - \sum_{V \in value(X)} \frac{|S_V|}{|S|} Entropy(S_V) \quad (2)$$

Keterangan:

- X adalah atribut
- V adalah suatu nilai yang mungkin untuk atribut (X)
- Value (X) adalah himpunan untuk atribut (X)

Setelah menghitung information gain pada semua atribut kemudian dicari information gain tertinggi untuk dijadikan root dalam suatu

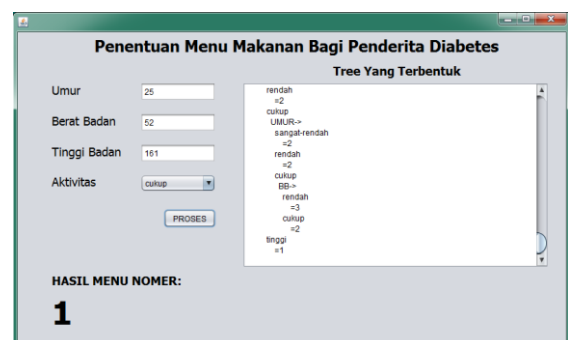
pohon keputusan. Begitu seterusnya sampai parameter tiap atribut terklasifikasi dengan sempurna.



Gambar 1. Diagram Alir ID3

5. IMPLEMENTASI

Implementasi antarmuka program selesai dilakukan ketika program telah selesai diimplementasikan pada bab sebelumnya. Gambar 2 menunjukkan hasil implementasi antarmuka program.



Gambar 2. Interface Program

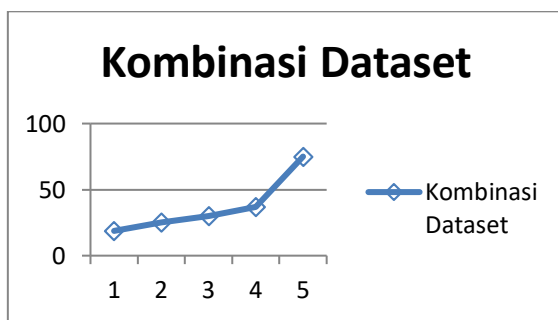
6. PENGUJIAN

Dataset yang akan digunakan sebanyak 100 data. Pengujian kombinasi data latih dan uji akan dilakukan untuk mengetahui tingkat akurasi terbesar dari sistem. Beberapa kombinasi digunakan untuk melakukan

pengujian dataset yang ditunjukkan hasilnya pada Tabel 1.

Tabel 1 Hasil pengujian kombinasi dataset

Percobaan	Kombinasi Data Latih : Data Uji (%)	Akurasi Testing
1	20:80	18,75%
2	40:60	25%
3	50:50	30%
4	60:40	37%
5	80:20	75%



Gambar 3 Hasil pengujian kombinasi dataset

Berdasarkan hasil pengujian kombinasi dataset kombinasi data latih dan data uji 80 dan 20 memberikan nilai akurasi rata-rata terbesar, yaitu 75%.

7. ANALISIS

Berdasarkan hasil pengujian dataset pada gambar 3, berikut merupakan pengujian menggunakan 20 data uji yang ditunjukkan pada Tabel 2.

Tabel 2 Hasil Pengujian Menggunakan 20 Data Uji

Umur	Parameter			Aktual	Hasil Sistem
	BB	TB	Aktivitas		
R	R	C	T	3	3
R	ST	C	C	3	3
R	T	R	ST	2	2
SR	C	R	ST	3	3
R	C	R	T	3	3
SR	C	C	ST	5	5
SR	C	C	ST	3	3
SR	R	R	ST	3	3
SR	C	R	T	3	4
SR	C	R	ST	3	3
SR	C	R	T	5	4
SR	C	C	ST	3	3
SR	C	R	T	2	2
SR	R	R	C	4	0
C	T	R	ST	5	0
C	T	R	ST	3	3

SR	C	R	T	3	3
SR	R	R	T	5	3
SR	C	C	T	5	5
SR	R	R	ST	3	3

Keterangan:

BB = Berat Badan

TB = Tinggi Badan

SR = Sangat Rendah

R = Rendah

C = Cukup

T = Tinggi

ST = Tinggi

Berdasarkan hasil pengujian menggunakan 20 dataset menunjukkan hasil akurasi sebesar 75 %. Akurasi yang dihasilkan masih tergolong cukup baik, hal ini karena ID3 memiliki beberapa kelemahan. Pertama ketika dataset kurang baik maka *tree* yang terbentuk tergolong kurang baik sehingga ketika dilakukan pengujian akan mendapatkan hasil yang kurang tepat. Kedua ID3 melakukan suatu keputusan berdasarkan *rule* yang terbentuk dari hasil pembentukan *tree*. Ketika ada beberapa data yang memiliki kesamaan dan hasil kelas yang berbeda maka hanya satu kelas yang akan terbentuk sehingga ketika diujikan hasil sistem tidak dapat mengklasifikasikan data yang akan karena tidak ada *rule*.

8. KESIMPULAN

Berdasarkan hasil penelitian diperoleh kesimpulan untuk menjalan rumusan masalah, sebagai berikut:

1. Perancangan sistem penentuan menu makanan yang aman bagi penderita penyakit diabetes mellitus dengan menggunakan metode ID3 dapat di implementasikan dengan baik. Metode penentu keputusan ID3 diformulasikan untuk menentukan menu makanan bagi penderita diabetes dari *tree* yang sudah terbentuk dari data *tranning* yang telah diproses
2. Berdasarkan hasil pengujian sistem penentuan menu makanan bagi penderita penyakit diabetes mellitus dengan metode ID3. Sedangkan akurasi terbaik pada saat pengujian kombinasi data *tranning* dan *testing* pada jumlah data *tranning* 80 data dan jumlah data *testing* 20 data diperoleh

akurasi data sebesar 75%. Beberapa hal yang menyebabkan hasil akurasi rendah antara lain:

- a. Dataset *training* sistem kurang baik maka *tree* yang terbentuk tergolong kurang baik sehingga ketika dilakukan pengujian dengan data baru akan mendapatkan hasil yang tidak sesuai.
- b. Metode ID3 melakukan suatu keputusan berdasarkan *rule* yang terbentuk dari hasil pembentukan *tree*. Ketika ada beberapa data yang memiliki kesamaan dan hasil kelas yang berbeda maka hanya satu kelas yang akan terbentuk sehingga ketika diujikan hasil sistem tidak dapat mengklasifikasikan data yang akan karena tidak ada *rule*.

Berdasarkan hasil penelitian didapatkan suatu simpulan bahwa metode ID3 belum maksimal untuk mengatasi pengklasifikasian menu makanan sehingga diperlukan penambahan metode untuk perbaikan *rule* yang akan dibentuk. Selain itu pada saat pelatihan diperlukan dataset yang banyak dan variatif sehingga dapat membentuk *rule* yang diperlukan.

9. DAFTAR PUSTAKA

- Dao, S. D. & Marian, R. 2011. Optimisation of precedence-constrained production sequencing and scheduling using genetic algorithms. *Proceedings of the International Multi Conference of Engineers and Computer Scientists*, 16-18 March, Hong Kong.
- Gen, M. & Cheng, R. 2000. *Genetic Algorithms and Engineering Optimization*. John Wiley & Sons, Inc., New York.
- Liliana, D. Y. & Mahmudy, W. F. 2006. Penerapan Algoritma Genetika pada Otomatisasi Penjadwalan Kuliah. *Laporan Penelitian DPP/SPP*. FMIPA Universitas Brawijaya, Malang.
- Marian, R. M., Luong, L. & Dao, S. D. 2012. Hybrid genetic algorithm optimisation of distribution networks—a comparative study. *Dalam: AO, S. I., CASTILLO, O. & HUANG, X. (editor.) Intelligent Control and Innovative Computing*. Springer, US.
- Phanden, R. K., Jain, A. & Verma, R. 2013. An approach for integration of process planning and scheduling. *International Journal of Computer Integrated Manufacturing*, 26(4), 284-302.
- Ridok, A. 2014. Peringkasan dokumen Bahasa Indonesia berbasis non-negative matrix factorization. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 1(1), 39-44.
- Tala, F. Z. 2003. A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. *Ph.D. Thesis*. Universiteit van Amsterdam.
- Wang, L. 2007. *Process planning and scheduling for distributed manufacturing*. Springer, London.
- Wibawa, A. P., Nafalski, A. & Mahmudy, W. F. 2013. Javanese `speech levels machine translation: improved parallel text alignment based on impossible pair limitation. *IEEE International Conference on Computational Intelligence and Cybernetics*, 3-4 December, Yogyakarta, Indonesia. 16-20.