



Penerapan Data Mining untuk Identifikasi Penyakit Diabetes Melitus dengan Menggunakan Metode Klasifikasi

Faizal Aris¹, Benyamin²

¹Dosen Program Studi Sistem Komputer, ²Program Studi Sistem Komputer, Stmik Bina Bangsa Kendari

¹Faizaries8@gmail.com, ²Benyamin@student.handayani.ac.id

Abstrak

Penyakit Diabetes Melitus (DM) dengan komplikasi merupakan penyebab tertinggi kematian ketiga di Indonesia yang setiap tahun penderitanya semakin bertambah, penyakit ini dulunya di juluki penyakit orang kaya namun seiring bertambahnya waktu penyakit ini sudah diidap oleh masyarakat menengah dan miskin. Hal ini dikarenakan bukan lagi karena faktor genetic tapi pola hidup yang tidak teratur menjadi penyumbang pesatnya penyakit ini, berdasarkan data WHO 80% penderita DM dapat dicegah, Klasifikasi pada penelitian ini bertujuan untuk memudahkan perawat dan penderita mengenali tipe penyakit DM agar penanganan penyakit diabetes semakin mudah dilakukan. Untuk menghasilkan informasi baru maka digunakan perhitungan algoritma C.45 dan pengujian algoritma yang menggunakan aplikasi rapid miner akan semakin memperkuat keputusan. Pada pengujian penelitian ini menggunakan beberapa atribut klasifikasi yakni atribut Jenis Kelamin, berat badan,Usia, Perokok, kadar gula darah, dan Tipe penyakit diabetes. Semua atribut tersebut akan dijadikan acuan dalam penelusuran hasil sehingga perawat dan penderita dapat menjadikan acuan dalam perawatan diri pasien secara optimal.

Kata Kunci : (data mining, diabetes mellitus, metode klasifikasi C.45)

Abstract

Diabetes mellitus (DM) with complications is the third highest cause of death in Indonesia, which every year the sufferer is increasing, this disease was once dubbed the disease of the rich, but over time the disease has been suffered by middle and poor communities. this is because it is no longer due to genetic factors but irregular lifestyles are contributors to the rapidity of this disease, based on WHO data 80% of DM patients can be prevented, the classification in this study aims to make it easier for nurses and sufferers to recognize the type of diabetes so that the handling of diabetes is easier do. To generate new information, the calculation of the C.45 algorithm and testing of the algorithm using the rapid miner application will further strengthen the decision. In testing this study uses several classification attributes, namely the Gender, weight, age, Smoker, blood sugar levels, and a history of diabetes. All of these attributes will be used as a reference in tracing the results so that Nurses and sufferers can make reference in self-care optimum Patients.

Keyword : (ata mining, diabetes mellitus, metode klasifikasi C.45)

1. Pendahuluan

Data mining merupakan gabungan sejumlah disiplin ilmu computer yang didefinisikan sebagai penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan irisan dari *artificial intelligence*, *machine learning*, *statistics*, dan *data base systems*. Data mining memiliki banyak sekali manfaat dalam pengolahan data sehingga data-data yang ada pada masa lampau dapat dijadikan kembali sebagai sumber pengetahuan dan informasi yang baru. Pemanfaatan data mining tidak hanya sebatas ilmu teknologi, tetapi juga pada bidang kesehatan. Data mining dapat dijadikan sebagai acuan untuk memprediksi dan mendiagnosa suatu jenis penyakit dengan menggunakan metode-metode yang dapat diterapkan. Dari sekian banyak penyakit, salah satu penyakit degeneratif yang dapat diprediksi dengan menggunakan metode data mining adalah penyakit diabetes. Penyakit diabetes atau yang sering dikenal dengan sebutan kencing manis ini merupakan penyakit di mana kadar glukosa (gula sederhana) di dalam darah menjadi tinggi karena tubuh tidak dapat memproduksi atau mengeluarkan insulin secara cukup. Penyakit diabetes dapat disebabkan oleh penderita yang memiliki riwayat penyakit diabetes turunan yang kita sebut Tipe Adan karena factor pemicu lainnya yang tidak berkaitan dengan factor turunan yang kita sebut Tipe B.

Diabetes mellitus (DM) merupakan kelainan heterogen yang ditandai dengan meningkatnya kadar glukosa dalam darah (hiperglikemia). DM merupakan salah satu penyakit metabolic kronik, dan jika tidak dilakukan pengobatan dan perawatan yang tepat dapat mengakibatkan kondisi yang membahayakan bahkan dapat menyebabkan komplikasi. Pada pasien dengan DM tipe B terjadi resistensi insulin atau gangguan sekresi insulin sehingga menyebabkan peningkatan kadar glukosa dalam darah atau hipergli kemia (Smeltzer & Bare, 2002).

Menurut International Diabetes Federation ([IDF], 2015) terdapat 415 juta orang mengalami diabetes pada tahun 2015 dan tahun 2040 diperkirakan akan meningkat mencapai 642 jutaorang. Jumlah DM setiap Negara meningkat dan usia terbanyak orang dengan DM berada di usia antara 40-59 tahun. DM menyebabkan kematian 5 juta jiwa pada tahun 2015. Data Riset Kesehatan Dasar (Riskesdas, 2013) menunjukkan jumlah kasus DM di Indonesia tahun 2010 sebesar 8,43 juta orang dan diperkirakan akan meningkat mencapai 21,25 juta di tahun 2035. Pada tahun 2013 jumlah DM di Indonesia dengan usia di atas 15 tahun sebesar 6,9%.

Jumlah penderita DM di Puskesmas Puwatu masuk dalam 10 besar penyakit yang diderita masyarakat pada tahun 2018 sebanyak 618 pasien.

NO.	JENIS PENYAKIT	JML
1.	Hipertensi	1419
2.	Kelainan Otot	955
3	ISPA	754
4.	Diabetes Melitus	618
5.	Infeksi Usus	543
6.	Kelainan Susunan Saraf	530
7.	Kelainan Jantung	462
8.	Penyakit Kulit	317
9	ASMA	198
10	Penyakit Pada Saluran Kemih	125
TOTAL		5921

Sumber: Data Primer, Puskesmas Puuwatu Tahun 2018.

Berdasarkan data diatas penulis tertarik untuk membuat sebuah klasifikasi identifikasi penyakit DM menggunakan metode klasifikasi algoritma C.45. Selain untuk menjadikan suatu keputusan kompleks menjadi lebih mudah, kelebihan lain yang didapatkan dari penelitian ini adalah pengujian sampel masalah tersebut hanya berdasarkan pada kriteria atau kelas tertentu. Penulis juga ingin mengimplementasikan metode tersebut dengan menggunakan aplikasi rapid miner agar dapat menghasilkan keputusan prediksi yang lebih akurat.

2. Telaah Pustaka

Penyakit *diabetes mellitus* dibagi menjadi 2 jenis yaitu:

1. *Diabetesmellitus*Tipe1 (tipe A)

Penyakit *diabetes mellitus* tipe 1 biasanya disebut *insulin dependent*. *Diabetes mellitus* tipe 1 ini terjadi pada usia muda dibawah 30 tahun. Seseorang yang menderita *diabetes mellitus* tipe 1 perlu dilakukan suntik insulin. Suntik insulin dilakukan karena glukosa darah dalam tubuh tidak dapat memproduksi insulin sebagaimana mestinya (Sa'dietal., 2015).

2. *Diabetes mellitus* Tipe 2 (tipe B)

Penyakit *diabetes mellitus* tipe 2 biasanya disebut *non-insulin dependent* yang ditandai dengan resistensi insulin dan gangguan sekresi insulin. Tipe ini sering diderita oleh seseorang yang berusia diatas 40 tahun. Hal ini terjadi ketika tubuh manusia tidak dapat secara aktif menggunakan insulin yang dihasilkan oleh tubuh. Biasanya disebabkan faktor keturunan, obesitas, kurangaktivitas, penyakit lain dan usia (V & Ravikumar, 2014).

Faktor Penyakit *Diabetes Mellitus*

Faktor risiko penyakit *diabetes mellitus* adalah suatu kondisi dimana kesehatan pada seseorang terkena penyakit *diabetes mellitus*. Apabila kondisi ini tidak ada penanganan khusus maka dapat memperburuk keadaan dan dapat mengakibatkan terjadinya penyakit komplikasi ataupun kematian. Terdapat beberapa faktor risiko pada penyakit *diabetes mellitus* seperti (Devi & Shyla, 2016):

1. Nafsu Makan Meningkat
2. Sering Buang Air Kecil
3. Peningkatan Kehausan
4. Turunnya Berat Badan
5. Usia (15-40) tahun
6. Faktor Keturunan
7. Mulut Kering
8. Mudah Kelelahan/Kurangnya Aktivitas Fisik

9. Sering Mengantuk

10. Mual/Muntah-Muntah

11. Timbulnya Luka yang Tak Kunjung Sembuh

12. Gatal-Gatal

13. Mengonsumsi makanan Berkolesterol Tinggi

14. Obesitas

15. Kadar Glukosa Darah Meningkat

Data mining

Data mining merupakan proses pengambilan data dari *data warehouse* berdasarkan prediksi variabel. *Data mining* juga sebagai penemuan pada KDD *database*. Prediksi digunakan untuk menemukan data dari sekumpulan data yang ditentukan dengan berbagai domain seperti kecerdasan buatan, basisdata, dan lain-lain. Penggunaan *data mining* memiliki dampak diberbagai bidang seperti bidang kesehatan. *Data mining* pada system medis dibutuhkan untuk mengekstrak informasi dari *data base* sehingga dapat melakukan diagnosis penyakit (Poonguzhali, Kabilan, Kannan, & Sivagami, 2014).

Metode Classification

Classification adalah Sebuah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Hal ini juga dapat dikatakan sebagai pembelajaran (klasifikasi) yang memetakan sebuah unsure (item) data kedalam salah satu dari beberapa kelas yang sudah didefinisikan (Bustami, 2013) [2].

Klasifikasi banyak digunakan dalam berbagai aplikasi, diantaranya adalah untuk mendeteksi kecurangan (fraud detection), pengelolaan pelanggan, diagnosis medis, prediksi penjualan dan lain sebagainya.

Algoritma C.45

Algoritma C45 merupakan metode yang menjadi pilihan pertama dan sering digunakan dalam pengembangan Data Mining karena kecepatan dalam pengklasifikasian pohon keputusan disamping dapat mengkonstruksi pengklasifikasian dengan aturan-aturan yang lain (Wu Xindong et al, 2008) [9].

Algoritma C4.5 merupakan kelompok algoritma Decision Tree. Algoritma ini mempunyai input berupa training samples dan samples. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data

Rapid Miner

Rapid miner merupakan *platform* perangkat lunak ilmu data yang dikembangkan oleh perusahaan yang bernama sama dengan pihak yang menyediakan lingkungan terintegrasi untuk persiapan data, penambangan teks, dan analisis prediktif. *Rapid miner* dilengkapi dengan satu prosesor logika dan 10.000 baris data. *Rapid miner* yang tersedia berada di bawah lisensi AGPL.

Diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau-) dari sejumlah data acak pada suatu ruang sampels. Entropy bisa dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai entropy maka semakin baik digunakan dalam mengekstraksi suatu kelas. Pada algoritma ID3 pengurangan entropy disebut dengan informasi gain. Pembagian sample S terhadap atribut A dapat dihitung information gain dengan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{v \in nilai(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Pilih atribut yang memiliki nilai information gain terbesar, ulangi proses perhitungan information gain akan terus dilaksanakan sampai semua data telah masuk dalam kelas yang sama. Atribut yang telah dipilih tidak diikuti lagi dalam perhitungan nilai information gain (Kristanto Obbie, 2011) [6].

3. Jenis dan Sumber Data

Data yang digunakan adalah data pasien penderita diabetes mellitus yakni *atribut jenis kelamin, berat badan, usia, kadar gula darah, dan tipe penyakit diabetes.*

4. Metode Penelitian

Berisi mengenai tahapan - tahapan penelitian dimulai dari penelitian awal berupa pengumpulan data sampai kepada penerapan algoritma kepada aplikasi *rapid miner* sehingga ditemukan kecocokan antara hasil akhir terhadap masing-masing perhitungan. *Output* dari penelitian ini adalah analisa terhadap hasil prediksi factor seseorang bisa terkena penyakit diabetes tipe A atau Tipe B. Pada penelitian ini, penulis menemukan beberapa keadaan yang membuat peneliti tertarik untuk melakukan suatu penganalisaan terhadap kemungkinan - kemungkinan yang akan membuat seseorang terkena penyakit diabetes. Seperti yang peneliti ketahui, penyakit diabetes tidak hanya dapat dijumpai pada orang yang memiliki keturunan penyakit diabetes, namun juga berpotensi pada seseorang yang memiliki gaya hidup tidak sehat serta beberapa faktor-faktor pemicu lainnya. Tahapan dari penelitian ini dilakukan dengan beberapa metode tahapan yang saling berhubungan antara satu dengan yang lainnya. Metode yang dilakukan berupa:

1. Mengumpulkan data pasien penderita diabetes. Pada tahapan ini dilakukan pengumpulan data dari pasien penderita diabetes. Pengumpulan data dilakukan melalui wawancara serta meminta sebagian

data pasien dari petugas puskesmas puwatu di Kendari.

2. Pengklasifikasian data – data pasien penderita diabetes. Setelah data–data dikumpulkan, maka dilanjutkan dengan proses mengelompokkan data ke dalam MS.Excel.
3. Melakukan pengolahan data dengan menggunakan software rapid miner.
4. Setelah data diklasifikasikan ke dalam Ms.Excel, maka data tersebut kemudian akan diproses dengan menggunakan software rapid miner. Dengan menggunakan software tersebut maka data akan dapat dianalisa untuk mendapatkan sebuah keputusan.

5. Hasil dan Pembahasan

Berikut merupakan data-data pasien penderita diabetes yang telah diklasifikasikan berdasarkan beberapa atribut yang sudah ditetapkan. Atribut-atribut tersebut terdiri dari berat badan, jenis kelamin, tekanan darah, kadargula darah dan tipe penyakit diabetes.

R1: IF glu sewaktu Normal AND TD Tinggi AND Usia >55 AND BB <=50 THEN B

R2: IF glu sewaktu Normal AND TD Tinggi AND Usia >55 AND BB 51-60 THEN A

R3: IF glu sewaktu Normal AND TD Tinggi AND Usia <=55 THEN B

7. Kesimpulan

Kesimpulan yang diperoleh dari penelitian ini adalah klasifikasi data penderita diabetes dengan teknik data mining klasifikasi yang menggunakan algoritma C.45 menghasilkan *rule* yang dapat digunakan untuk prediksi penyakit diabetes. Hasil dari pohon keputusan berisi tentang analisa terhadap faktor – faktor seseorang berpotensi terkena penyakit diabetes dengan melihat kepada atribut – atribut seperti jenis

R4: IF glu sewaktu Normal AND TD Normal THEN B

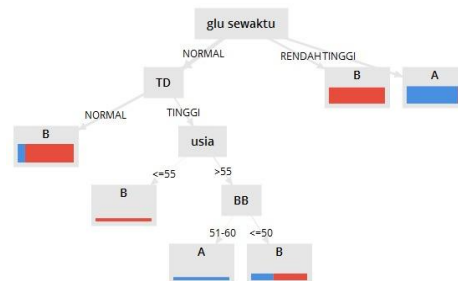
R5: IF glu sewaktu Rendah THEN B

R6: IF glu sewaktu Tinggi THEN A

Berat badan <=50, normal, 51-60 berat dan >60 obesitas

6. Pengujian Hasil

Sistem aplikasi yang akan peneliti gunakan untuk pengolahan data pasien dengan resiko diabetes adalah *rapidminer* 9.1 Berikut adalah gambar hasil pohon keputusan yang didapatkan setelah data dikelola dengan menggunakan rapid miner 9.1.



Gambar. Pohon keputusan menggunakan rapid miner 9.1

kelamin (jenis kelamin hanya sebagai atribut bantuan yang tidak dapat dijadikan perhitungan dalam memprediksi riwayat penyakit diabetes), berat badan, tekanan darah, dan kadar gula darah dan variabel riwayat penyakit turunan dan tidak turunan.

Saran untuk penelitian selanjutnya adalah komparasi dengan beberapa algoritma untuk prediksi penyakit diabetes sehingga diperoleh algoritma dengan tingkat akurasi tertinggi.



Selain itu penelitian dilakukan dengan data diabetes yang lain dengan jumlah yang lebih banyak.

8. Daftar Pustaka:

- Anik andriani, sistem prediksi penyakit diabetes berbasis decision tree, Jurnal bianglala informatika, vol 1 no 1 september 2013
- Devi,M.R.,& Shyla,J.M. (2016). Analysisof Various Data Mining Techniques to Predict Diabetes Mellitus. *International Journal of Applied Engineering Research*, 11, 727–730.
- Jerhi wahyu Fernanda, boosting neural network dan boosting chart pada klasifikasi diabetes mellitus tipe II, jurnal matematika vol 2 no.2, desember 2012 ISBN: 1693-1394
- Okriandi nugroho, bagus adhi kusuma, zanuar rifai, tri astute, uswatu khasanah, risky wahyudi, peranan data mining untuk mengidentifikasi penyakit diabetes mellitus dengan menggunakan algoritme ID3, ISBN: 978-602-60280-1-3
- Poonguzhali,E., Kabilan,S.,Kannan,S., & Sivagami, P.(2014). Diagnosisof Diabetes Mellitus Type2using NeuralNetwork. *International Journal of Emerging Technology and Advanced Engineering*,4(2),939–942.
- Rima Ulfa Fahra, NurWidayati, Jon Hafan Sutawardana (2017). Hubungan Peran Perawat Sebagai Edukator Dengan Perawatan Diri Pasien Diabetes Melitus Tipe 2 Di Poli Penyakit Dalam Rumah Sakit Bina Sehat Jember
- Silvia ikmalia Fernanda, dian eka ratnawati, putra pandu adikara, identifikasi penyakit diabetes mellitus dengan menggunakan metode modified K-Nearest Neighbor (MKNN), jurnal pengembangan teknologi informasi dan ilmu computer, vol 1 no. 6
- Smeltzer, S.C. & Bare, B.G.(Eds) 2002. Buku Ajar Keperawatan Medikal-Bedah Brunner & Suddarth. Ed.8,Vol.2. Jakarta: EGC
- Sa'di,S.,Maleki, A., Hashemi,R., Panbechi,Z., & Chalabi, K.(2015). Comparison of Data Mining Algorithmsin the Diagnosisof TypeIiDiabetes. In*International Journal on Computational Science & Applications* (Vol. 5,hal. 1–12).
- V.,&Ravikumar,A.(2014). Study of Data Mining Algorithms for Prediction and Diagnosis of Diabetes Mellitus. *International Journal of Computer Application*,95(17), 12–16.