

PENERAPAN ENTROPY BASED DISCRETIZATION PADA METODE NAIVE BAYES DALAM KLASIFIKASI PENYAKIT DIABETES MELLITUS

Rudy Ariyanto¹, Dwi Puspitasari², Yusniar Alfani³

^{1,2,3} Jurusan Teknologi Informasi, Politeknik Negeri Malang

¹ariyantorudy@polinema.ac.id, ²dwi.puspitasari@polinema.ac.id, ³yusniaralfani10@gmail.com

Abstrak

Diabetes Mellitus merupakan salah satu penyakit penyebab kematian tertinggi di Indonesia. Penderita Diabetes Mellitus diperkirakan akan terus meningkat. Hal ini tersebut menunjukkan bahwa Diabetes Mellitus perlu diperhatikan secara serius. Penyakit diabetes dapat dicegah jika kita mengetahui gejala – gejala penyakit Diabetes Mellitus. Dari permasalahan di atas, maka dibuatlah sistem yang berjudul “Penerapan Entropy Based Discretization pada Metode Naive Bayes Dalam Klasifikasi Penyakit Diabetes Mellitus”. Sistem ini membantu sarana kesehatan Puskesmas Kendalsari dalam melakukan anamnesa dan menentukan hasil tes laboratorium. Petugas anamnesa akan melakukan anamnesa dengan pasien dengan keluaran tingkat potensi user terhadap penyakit diabetes. Tingkat potensi pasien terhadap Diabetes Mellitus dikelompokkan menjadi tiga kelas yaitu, rendah, sedang dan tinggi. Selanjutnya petugas laboratorium akan pemeriksaan terhadap pasien untuk memastikan apakah hasil pemeriksaan sesuai dengan tingkat potensi user terhadap penyakit Diabetes Mellitus. Metode Naive Bayes adalah metode yang digunakan melakukan klasifikasi potensi user dan hasil laboratorium pasien. Sedangkan Metode Entropy Based Discretization digunakan untuk melakukan transformasi hasil laboratorium pasien yang memiliki nilai kontinu menjadi bernilai diskrit. Data gejala Diabetes Mellitus dalam penelitian ini diambil dengan menggunakan kuisioner, sedangkan data hasil tes dalam penelitian ini diambil dari Puskesmas Kendalsari. Dari hasil implementasi klasifikasi gejala user menggunakan metode Naive Bayes mendapatkan akurasi sebesar 96%. Sementara hasil implementasi klasifikasi hasil tes user menggunakan metode transformasi Entropy Based Discretization mendapatkan akurasi sebesar 80%. Saran untuk sistem ini adalah penambahan jadwal kontrol user kepada dokter agar gula darah selalu terjaga.

Kata kunci : *Diabetes Mellitus, Naive Bayes, Entropy Based Discretization*

1. Pendahuluan

Teknologi informasi berkembang sejalan dengan perkembangan peradaban manusia. Saat ini perkembangan teknologi informasi juga telah diterapkan pada berbagai bidang termasuk bidang kesehatan. Pada bidang kesehatan, salah satu penyakit yang menjadi beban utama adalah penyakit kronis dengan ciri bersifat menetap, menyebabkan ketidakmampuan pada penderitanya, dan untuk menyembuhkannya penderita perlu melakukan perawatan dalam periode waktu yang lama. Salah satu penyakit kronis yang menjadi prioritas di bidang kesehatan adalah *Diabetes Mellitus*.

Diabetes Mellitus merupakan salah satu penyakit penyebab kematian tertinggi di Indonesia. Diabetes Mellitus juga merupakan penyebab utama untuk kebutaan, serangan jantung, stroke, gagal ginjal dan amputasi kaki. Penderita Diabetes Mellitus diperkirakan akan terus meningkat. World Health Organization memperkirakan jumlah penderita diabetes di Indonesia akan meningkat

signifikan hingga 21,3 juta jiwa pada 2030 mendatang. Tingginya angka tersebut menunjukkan bahwa Diabetes Mellitus perlu diperhatikan secara serius. Penyakit diabetes dapat dicegah jika kita mengetahui anamnesa – anamnesa penyakit Diabetes Mellitus. Pencegahan terhadap penyakit Diabetes Mellitus dapat dilakukan dengan cara menerapkan sistem pakar yang merupakan ilmu komputer yang mengadopsi pengetahuan ahli ke dalam komputer agar komputer dapat menyelesaikan masalah seperti yang biasa dilakukan oleh ahli. Sistem pakar ini akan membantu kita dalam memberikan jawaban dari pertanyaan seputar diabetes Mellitus.

Dari permasalahan di atas, maka dibuatlah sistem yang berjudul “Penerapan *Entropy Based Discretization* pada Metode *Naive Bayes Classifier* Dalam Klasifikasi Penyakit Diabetes Mellitus”. Sistem ini diharapkan dapat membantu salah satu sarana kesehatan yaitu, Puskesmas Kendalsari dalam mengetahui tingkat potensi pasien terhadap penyakit Diabetes Mellitus. Hal ini biasa dikenal dengan skrining, suatu proses dimana pasien akan menjawab

pertanyaan anamnesa dari petugas anamnesa terkait dengan diabetes Mellitus atau gaya hidup pasien. Hasil keluaran yang diberikan oleh sistem pakar, dapat dijadikan sebagai diagnosa awal dan dilanjutkan dengan pemeriksaan tes laboratorium untuk mengetahui hasil yang lebih pasti tentang status pasien terhadap penyakit Diabetes Mellitus. Keluaran tingkat potensi pasien terhadap penyakit Diabetes Mellitus terbagi menjadi tiga yaitu, rendah, sedang dan tinggi. Keluaran tersebut diperoleh dari pertanyaan-pertanyaan terkait dengan anamnesa yang sering dialami atau gaya hidup penderita Diabetes Mellitus. Sistem ini juga membantu dalam untuk dengan memasukkan hasil pemeriksaan dari laboratorium dengan keluaran yaitu, normal, pradiabetes dan diabetes.

Metode *Naive Bayes* digunakan untuk melakukan klasifikasi hasil *anamnesa* Diabetes Mellitus user. Karakteristik dari metode ini adalah tiap fitur atribut data dianggap independen, satu dan lainnya terpisah dan memiliki nilai sendiri. Metode ini dianggap memiliki performa yang handal dan kompetitif dalam proses pengklasifikasian karena asumsi independen atribut yang dimiliki sebuah data sangat sesuai dengan pengaplikasian di dunia nyata. Sedangkan Metode *Entropy Based Discretization* digunakan untuk transformasi pada atribut anamnesa dan hasil tes pasien yang memiliki nilai kontinyu. Data anamnesa dan data hasil tes dalam penelitian ini diambil dari Puskesmas Kendalsari.

2. Landasan Teori

2.1 Diabetes Mellitus

Diabetes Mellitus adalah keadaan hiperglikemia kronik disertai berbagai kelainan metabolik akibat gangguan hormonal, yang menimbulkan berbagai komplikasi kronik pada mata, ginjal, dan pembuluh darah, disertai lesi pada membran basalis dalam pemeriksaan dengan mikroskop elektron. Penyakit diabetes sering mengancam kesehatan semua orang di seluruh dunia. Penyakit diabetes disebut juga penyakit kompliasi karena banyak anamnesa yang dapat mengakibatkan penyakit lainnya ataupun kematian.

Adapun beberapa pertanyaan yang akan dijawab oleh pasien melalui petugas anamnesa adalah

1. Umur
2. Lingkar Pinggang
3. Body Mass Index
4. Apakah pasien menghabiskan waktu setidaknya 30 menit setiap hari untuk berolahraga?
 - a. Ya
 - b. Tidak
5. Seberapa sering pasien mengonsumsi sayur?
 - a. Setiap hari
 - b. Tidak setiap hari

6. Apakah pasien pernah melakukan perawatan atau minum obat untuk gula darah tinggi?
 - a. Ya
 - b. Tidak
7. Apakah pasien pernah mengalami gula darah tinggi?
 - a. Ya
 - b. Tidak
8. Apakah keluarga pasien ada yang mengalami Diabetes Mellitus?
 - a. Tidak ada
 - b. Ya, kakek/nenek, tante/paman,
 - c. Ya, orang tua, kakak/adik

Untuk mengetahui hasil tes pemeriksaan laboratorium terdapat beberapa atribut yang harus diisi oleh petugas lab, diantaranya

Tabel 1. Atribut Hasil tes

No	Atribut
1.	Jenis kelamin
2.	Umur
3.	Gula darah puasa
4.	Gula darah 2 jam sesudah makan

Berikut adalah *sample* data yang digunakan dalam penelitian ini

Tabel 2. *Sample* data

No	JK	Age	BMI	LP	Olahraga	Sayur	Medicator/Blood Gl	Keturunan Keluarga	Class	N	Danj	Hasil Tes	
1	1	37	24	92	2	2	1	1	1	1	90	120	1
2	1	45	23	92	2	1	1	1	1	1	95	110	1
3	1	44	24	93	2	1	1	1	2	1	88	110	1
4	1	30	22	92	1	1	1	1	1	1	91	105	1
5	2	48	25	82	2	2	1	1	1	1	90	120	1
6	1	42	23	93	2	1	2	2	1	1	90	145	2
7	2	35	22	76	1	1	1	1	3	1	94	111	1
8	2	32	31	90	2	1	1	1	1	1	128	203	3
9	1	56	24	93	2	2	1	1	2	1	80	147	2
10	2	53	22	76	2	1	1	1	2	1	85	100	1

2.2 Naive Bayes

Naive Bayes adalah metode pengklasifikasian probabilitas sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang ada. *Naive Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai *output*.

Naive Bayes digunakan untuk melakukan klasifikasi terhadap data anamnesa pasien. *Naive Bayes* pada proses training data anamnesa pasien, akan menghitung jumlah probabilitas kemunculan suatu angka pada tiap atribut. Lalu dijumlahkan dan mencari nilai terbesar dan menjadikan nilai terbesar tersebut sebagai hasil akhir. Keluaran dari *Naive Bayes* ini adalah potensi user terhadap penyakit Diabetes Mellitus, yaitu rendah, sedang dan normal.

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Dimana:

X : Data dengan *class* yang belum diketahui

H : Hipotesis data

$P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X

$P(H)$: Probabilitas hipotesis H

$P(X|H)$: Probabilitas X berdasarkan kondisi hipotesis H

$P(X)$: Probabilitas X

Langkah – langkah penyelesaian metode ini adalah

- Menghitung $P(X)$ untuk setiap kelas
- Menghitung $P(X|H)$ untuk setiap kriteria dari setiap kelas
- Mencari nilai $P(X|H)$ yang paling besar dan menjadikannya sebagai hasil akhir

2.3 Entropy Based Discretization

Entropy Based Discretization adalah metode *supervised discretization* yang menggunakan informasi kelas entropy dalam melakukan diskretisasi. Metode ini menentukan *information gain* terkecil dari semua atribut yang ingin ditransformasi dan menjadikannya sebagai *split point*.

Naive Bayes adalah metode klasifikasi yang melakukan *training* dengan cara menghitung probabilitas kemunculan suatu nilai. Jika terdapat atribut yang bernilai kontinu maka kecil kemungkinan probabilitas atribut tersebut. Oleh karena itu, *Entropy Based Discretization* ini bertugas untuk melakukan transformasi atribut bernilai kontinu menjadi diskrit sehingga probabilitas kemunculan suatu nilai semakin besar. Di dalam penelitian ini, terdapat beberapa atribut bernilai kontinu seperti, umur, lingkar pinggang, dan body mass index. Atribut – atribut tersebut akan mengalami proses transformasi terlebih dahulu sebelum akhirnya memasuki proses klasifikasi.

Split point adalah acuan nilai dalam melakukan transformasi pada metode *Entropy Based Discretization*. *Split point* paling kecil lah yang akan dijadikan acuan karena dapat membagi interval menjadi rata. Nilai yang lebih kecil dari *split point* berpeluang untuk masuk ke dalam kelas yang baik sementara nilai yang lebih besar dari *split point* berpeluang untuk masuk ke dalam kelas yang kurang baik seperti berpotensi sedang/tinggi, pradiabetes dan diabetes.

Langkah – langkah penyelesaian metode ini sebagai berikut

- Menentukan $Entropy(S)$

$$Entropy(S) = - \sum p_i \log_2(p_i)$$

Dimana :

p_i : probabilitas kelas yang diinginkan, didapatkan dari C_i/S .

C_i adalah probabilitas kelas yang dicari
 S adalah probabilitas keseluruhan.

- Menentukan $Entropy(S,A)$

$$Entropy(S,A) = \sum \frac{|S_i|}{|S|} * Entropy(S_i)$$

Dimana :

$|S_i|$ = jumlah kasus pada partisi ke i

$|S|$ = jumlah kasus dalam himpunan S

A = atribut

n = jumlah partisi atribut A

- Menentukan *Gain*

$$Gain(S,A) = Entropy(S) - \sum \frac{|S_i|}{|S|} * Entropy(S_i)$$

- Menentukan *Info Gain(S,T)*

$$InfoGain(S,T) = Info(S) - Info(S,T)$$

Menjadikan *Info Gain* terkecil sebagai *split point*

2.4 Sistem Pakar

Sistem pakar adalah sistem yang berusaha mengadopsi pengetahuan manusia ke komputer, agar komputer dapat menyelesaikan masalah seperti yang biasa dilakukan oleh para ahli, dan sistem pakar yang baik dirancang agar dapat menyelesaikan suatu permasalahan tertentu dengan meniru kerja dari para ahli.

Pakar yang dimaksud disini adalah orang yang mempunyai keahlian khusus yang dapat menyelesaikan masalah yang tidak dapat diselesaikan orang awam. Contohnya dokter, mekanik, psikolog, dan lain-lain.

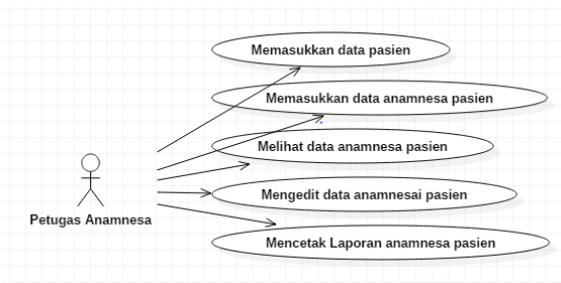
Dalam penelitian ini, pakar yang dimaksud adalah dokter. Sistem ini diharapkan mampu menyelesaikan masalah seperti seorang dokter yang dapat mendiagnosa tingkat potensi user terhadap Diabetes Mellitus dan dapat mengetahui hasil pemeriksaan laboratorium user.

3. Analisis dan Perancangan

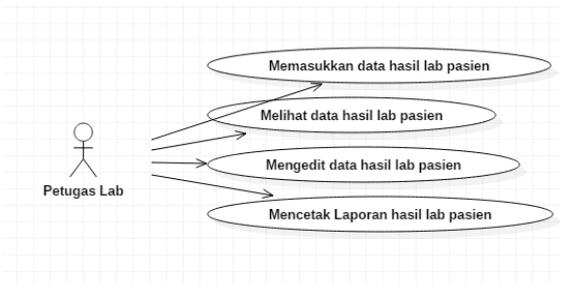
3.1 Use Case Diagram

Use case diagram merupakan diagram yang menggambarkan interaksi antara aktor dan inisiator dari interaksi sistem. Petugas anamnesa bertugas memasukkan data pasien beserta jawaban yang akan ditanyakan kepada pasien. Petugas anamnesa pertanyaan seperti umur, mengukur bmi, lp, intensitas pasien makan sayur dan olahraga, keturunan dan riwayat kesehatan. Kemudian Petugas anamnesa dapat mengetahui hasil dari jawaban – jawaban tersebut. Keluaran dari sistem ini yaitu potensi user terhadap penyakit Diabetes Mellitus.

Terdapat tiga keluaran potensi yaitu, rendah, sedang dan tinggi. Petugas anamnesa juga dapat melakukan edit sebelum akhirnya laporan anamnesa dapat dicetak. Petugas anamnesa dapat mencetak laporan hasil anamnesa. Selanjutnya pasien akan diarahkan menuju petugas lab untuk melakukan pemeriksaan laboratorium

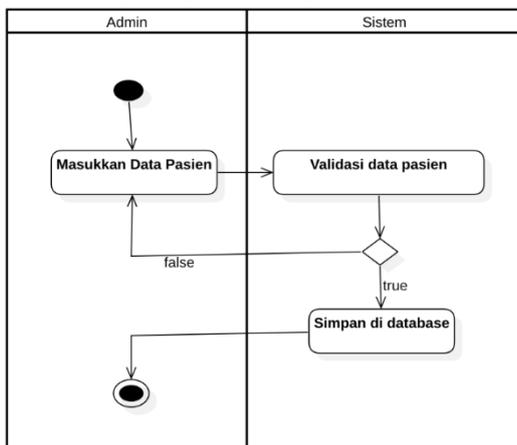


Gambar 1 usecase petugas anamnesa



Gambar 2 usecase petugas laboratorium

3.2 Activity Diagram Petugas Anamnesa



Gambar 3. Proses Penambahan Data Pasien

Petugas Anamnesa menambahkan data pasien sebelum melakukan anamnesa. Jika terdapat data yang belum diisi maka akan muncul pesan *error*

3.3 Pembahasan Naive Bayes

Berikut contoh data anamnesa pasien

- a. Hitung probabilitas tiap kelas
 - Kelas 1 = $2 / 7 = 0,28$
 - Kelas 2 = $3 / 7 = 0,42$
 - Kelas 3 = $2 / 7 = 0,28$
- b. Hitung probabilitas tiap *atribut* dibanding kelas
 - Atribut a1
 - $A1 (1) 1 = 3 / 2 = 1,5$
 - $A1 (1) 2 = 1 / 3 = 0,33$
 - $A1 (1) 3 = 0$
 - $A1 (2) 1 = 0$
 - $A1 (2) 2 = 1 / 3 = 0,33$

Tabel 3. Perhitungan

a1	a2	a3	Kelas
1	1	1	1
1	1	2	1
2	2	3	2
1	2	2	2
2	2	2	2
3	3	2	3
2	3	3	3

$A1 (2) 3 = 1 / 2 = 0,5$

- c. Lakukan proses *testing*
- d. Untuk data 1

A1 yang bernilai 1 maka, carilah probabilitas A1 yang bernilai 1 untuk semua kelas

4. Implementasi dan pengujian

Penelitian ini menghasilkan dua aplikasi yaitu aplikasi web dan mobile. Pada aplikasi web diimplementasikan pada computer sedangkan aplikasi *mobile* diimplementasikan pada Android *platform*. Aplikasi web ditujukan kepada admin untuk mengelola data-data yang ada sedangkan aplikasi *mobile* ditujukan kepada mahasiswa dan dosen.

4.1 Pengujian Akurasi

- Pengujian akurasi dengan rumus =

$$\frac{\text{jumlah data yang sesuai dengan data asli dan data hasil klasifikasi} * 100\%}{\text{jumlah data keseluruhan}}$$

$$= \frac{29}{30} * 100\%$$

$$= 96\%$$

- Pengujian di Puskesmas Kendalsari

Pengujian aplikasi dilakukan di Puskesmas Kendalsari dengan tujuan melakukan uji coba apakah hasil akhir dari aplikasi memiliki hasil yang sama dengan diagnosa dokter. Pengujian dilakukan terhadap 3 pasien dan 1 petugas anamnesa. Pasien menjawab pertanyaan yang ada pada aplikasi.

Gambar 4. Form Anamnesa

Kemudian petugas anamnesa memberitahukan kepada pasien hasil resiko atau potensi dan *preview* dari jawaban anamnesa pasien.

M. Saiful Anwar	Selorejo C 19	Laki-laki	46	27	95	Tidak setiap hari	Tidak setiap hari	Pemah	Pemah	Ya, ada ortu, kakak/adik	Tinggi	Hasil Tes
-----------------	---------------	-----------	----	----	----	-------------------	-------------------	-------	-------	--------------------------	--------	-----------

Gambar 5. Hasil Anamnesa

Setelah mengetahui hasil resiko atau potensi, pasien melakukan pemeriksaan laboratorium dan petugas memasukkan hasil pemeriksaan

Gambar 6. Form Hasil Tes

Kemudain petugas memberitahukan kepada pasien hasil tes

M. Saiful Anwar	Selorejo C 19	Laki-laki	46	27	95	Tidak setiap hari	Tidak setiap hari	Pemah	Pemah	Ya, ada ortu, kakak/adik	Tinggi	155	265	Diabetes
-----------------	---------------	-----------	----	----	----	-------------------	-------------------	-------	-------	--------------------------	--------	-----	-----	----------

Gambar 7. Hasil Anamnesa dan Hasil Tes

5. Kesimpulan dan saran

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan mengenai “Penerapan *Entropy Based Discretization* pada Metode *Naive Bayes* dalam Klasifikasi Penyakit *Diabetes Mellitus*” maka diambil kesimpulan bahwa:

1. Sistem ini berhasil menerapkan metode *Naive Bayes* sebagai klasifikasi dan metode *Entropy Based Discretization* sebagai metode transformasi dengan akurasi 96%.
2. Hasil pengujian menunjukkan bahwa sistem ini dapat membantu Puskesmas Kendalsari untuk mengetahui tingkat potensi user dan hasil tes pemeriksaan laboratorium terhadap penyakit *Diabetes Mellitus* dengan tingkat akurasi yang tinggi.

5.2 Saran

Sistem ini dapat ditambahkan jadwal kontrol *user* dengan dokter agar gula darah selalu terjaga.

Daftar Pustaka:

- Raghupathi, Wullianallur and Raghupathi, Viju, “An Empirical Study of Chronic Disease in the United States: A Visual Analytics Approach to Public Health”, J. Environ. Res. Public Health., March 2018
- World Health Organization. (2016). Diabetes Fakta dan Angka [online]. Available: <http://www.searo.who.int/indonesia/topics/8-whd2016-diabetes-facts-and-numbers-indonesian.pdf>.
- Skripsi Mahasiswa Sanata Dharma Yogyakarta. 2016 [online]. Available:

https://repository.usd.ac.id/9336/2/125314074_full.pdf.

Riadi, Muchlisin (20 Oktober 2016). Pengertian, Tujuan dan Struktur Sisten Pakar. Available: <https://www.kajianpustaka.com/2016/10/pengertian-tujuan-dan-struktur-sistem-pakar.html>

Rajashree, Rasmita and Rajib, “Comparative Analysis of Supervised and Unsupervised Discretization Techniques, vol, 2, no 3, April 2011

Sisodia Deepty, Sisodia Dhilip. “Prediction of Diabetes using Classification Algorithms ”, Proc. ICCIDS 2018

Nidhi, Mukesh, and Latika, “Classification of Diabetes patient by using Data Mining Techniques”, Vol-04, Issue-05, Aug 2018

