

KLASIFIKASI PENYAKIT DIABETES DENGAN HIDDEN NAIVE BAYES

Bagus Setya Rintyarna^{*1}

Abstrak

Penyakit Diabetes termasuk salah satu jenis penyakit yang perlu diwaspadai karena memiliki tingkat prevalensi yang cukup tinggi. Sebagai upaya deteksi dini penyakit Diabetes, pada penelitian ini digunakan Hidden Naïve Bayes sebagai metode untuk klasifikasi penyakit Diabetes. Hasil pengujian menunjukkan bahwa Hidden Naïve Bayes dapat digunakan untuk klasifikasi penyakit Diabetes dengan kinerja yang lebih baik dibandingkan Naïve Bayes Classifier.

Keywords : Hidden Naïve Bayes, Naïve Bayes Classifier

PENDAHULUAN

Diabetes termasuk ke dalam jenis/kelompok penyakit metabolik (metabolic disorder) yaitu suatu jenis penyakit yang berkaitan dengan produksi energi dalam sel manusia. Sebagian besar penyakit metabolik, termasuk diabetes adalah penyakit genetik atau penyakit keturunan dan sisanya disebabkan oleh racun, makanan dan infeksi. Karakteristik utama terjangkit penyakit diabetes adalah meningkatnya kadar gula dalam darah (hiperglikemia) yang disebabkan oleh terjadinya kelainan sekresi insulin, kerja insulin atau keduanya.

Terdapat tiga jenis penyakit diabetes yaitu Diabetes Melitus Tipe 1 di mana tubuh tidak memproduksi insulin sama sekali, Tipe ini adalah jenis diabetes yang tergantung kepada insulin dan biasanya muncul pada usia 40 tahun. Sampai saat ini diabetes jenis ini tidak dapat dicegah karena penyebabnya bukan pola makan yang kurang seimbang tetapi disebabkan oleh adanya kesalahan reaksi antibody pada tubuh penderita sehingga menghancurkan sel beta pancreas. Jenis kedua adalah Diabetes Melitus Tipe 2 di mana tubuh tidak cukup memproduksi insulin atau insulin tidak bekerja dengan seharusnya. Jenis ini berkembang ketika tubuh masih mampu menghasilkan insulin akan tetapi insulin yang dihasilkan tidak cukup atau bisa juga disebabkan oleh insulin resistance yang menyebabkan insulin tidak bisa bekerja sebagaimana mestinya. Sebagian besar penderita diabetes (90-95%) adalah penderita Diabetes Tipe 2. Dan tipe terakhir adalah Gestational Diabetes yang biasanya terjadi pada masa kehamilan.

Menurut laporan The International Diabetes Federation, pada tahun 2003 jumlah penduduk dunia yang terjangkit diabetes sebanyak 194 juta dan akan meningkat menjadi sekitar 333 juta pada tahun 2025 (Yoon, K.H., et al, 2006). Indonesia sendiri adalah Negara dengan peringkat ke-5 dengan jumlah penderita diabetes sebesar 12.9 juta jiwa dalam angka statistic penderita diabetes berdasarkan laporan dari World Health Organisation (WHO) pada tahun 2003. Dan diperkirakan akan naik menjadi peringkat ke-3 pada tahun 2025 dengan penderita diabetes sebanyak 20.9 juta jiwa.

Mengingat bahaya dan perkembangannya yang cepat baik di dunia, khususnya di Indonesia maka diperlukan langkah antisipasi yang tepat yang dilakukan oleh berbagai macam pihak dengan spesifikasi disiplin ilmunya masing-masing. Dalam bidang komputer, khususnya machine learning, diperlukan sebuah algoritma dengan tingkat akurasi dan presisi tinggi untuk mendeteksi penyakit diabetes. Implementasi algoritma tersebut diharapkan mampu memberikan bantuan dalam bentuk deteksi dini penyakit diabetes dengan tingkat akurasi dan presisi yang baik. Oleh karena itu pada usulan penelitian kali ini akan diuji performa klasifikasi algoritma Hidden Naïve Bayes untuk mendeteksi penyakit diabetes. Penelitian yang diusulkan berjudul "*Klasifikasi Penyakit Diabetes Menggunakan Algoritma Hidden Naïve Bayes*".

DIABETES MELLITUS

Diabetes adalah kelainan pada kemampuan tubuh untuk mengkonversi glukosa menjadi energy. Glukosa adalah sumber energi utama dalam tubuh manusia. Ketika makanan dicerna dalam tubuh, sumber makanan diubah menjadi lemak, protein atau karbohidrat. Jenis makanan yang berpengaruh terhadap kadar

gula darah adalah karbohidrat. Karbohidrat ketika dicerna diubah menjadi glukosa. Glukosa kemudian ditransfer dalam darah untuk dipergunakan oleh sel sebagai sumber energy. Agar glukosa bisa ditransfer dalam darah, diperlukan peran hormone dalam tubuh yaitu hormone insulin. Insulin diproduksi oleh sel beta dalam pancreas. Pada seseorang yang menderita diabetes, proses transfer ini mengalami kerusakan karena pankreas gagal memproduksi jumlah insulin yang cukup untuk mentransfer gula dalam darah.

Ada dua jenis penyakit diabetes, yaitu : 1) Diabetes Tipe 1 dan 2) Diabetes Tipe 2. Pada Diabetes Tipe 1 terjadi akibat faktor genetik dan sampai sekarang penyebabnya belum bisa ditentukan secara pasti. Prosentase penderita Diabetes Tipe 1 terhadap keseluruhan penderita penyakit diabetes berkisar antara 5-10%. Diabetes Tipe 1 adalah Diabetes yang tergantung kepada insulin di mana tubuh kekurangan hormone insulin yang dalam dunia medis dikenal dengan istilah Insulin Dependent Diabetes Mellitus (IDDM). Hal ini disebabkan karena hilangnya sel beta penghasil insulin pada pulau-pulau Langerhans pada pancreas. Diabetes Tipe 1 banyak ditemukan pada balita, anak-anak dan remaja. Sampai saat ini, Diabetes Tipe 1 hanya dapat diobati dengan pemberian terapi insulin yang dilakukan secara terus menerus berkesinambungan. Riwayat keluarga, diet dan faktor lingkungan sangat mempengaruhi perawatan penderita penyakit Diabetes Tipe 1.

Prosentase penderita paling banyak adalah Diabetes Tipe 2 yang berkorelasi sangat kuat dengan pola makan, obesitas dan aktivitas tubuh. Prosentase penderita Diabetes Tipe 2 terhadap keseluruhan penderita penyakit diabetes berkisar antara 90-95%. Tipe ini dikenal pula dengan istilah Non-Insulin Dependent Diabetes Mellitus (NIDDM). Pada penderita jenis ini, pengontrolan kadar gula darah dapat dilakukan dengan beberapa tindakan seperti diet, penurunan berat badan, dan pemberian tablet diabetik. Apabila pemberian tablet belum memberikan hasil yang optimal, perlu dipertimbangkan penggunaan obat suntik. Meskipun sebenarnya upaya pencegahan lebih perlu dilakukan untuk menanggulangi munculnya penyakit diabetes sedini mungkin. Dalam laporan penelitian berjudul "Prevention of Type 2 Diabetes Mellitus by Change in Lifestyle Among Subjects with Impaired Glucose Tolerance" yang melakukan pengujian terhadap 522 orang dengan resiko tinggi diabetes sebagai sampel uji selama kurang lebih 4 tahun didapatkan kesimpulan bahwa ternyata Penyakit Diabetes Tipe 2 dapat dicegah pada orang dengan resiko tinggi dengan cara mengubah life style (Tuomilehto, et al, 2001).

Gejala umum yang muncul pada penderita Diabetes antara lain adalah : banyak kencing (polyuria), haus dan banyak minum (polidipsia), lapar (polyfagia), letih dan lesu, penurunan berat badan yang tidak dapat dijelaskan sebabnya, lemah badan, kesemutan, gatal, pandangan gatal, disfungsi ereksi pada pria dan pruritus vulvae pada wanita.

Penanganan penyakit diabetes umum dilakukan dengan cara bertahap mulai dengan mengubah gaya hidup khususnya dalam pola makan, kemudian bila tidak menunjukkan hasil yang baik dilakukan dengan pemberian obat oral. Berdasarkan evaluasi tingkat keberhasilannya maka dosis obat dinaikkan dan jika pada akhirnya obat oral juga tidak memperbaiki keadaan maka kemudian diberikan suntikan insulin. Tujuan pemberian insulin adalah untuk mengontrol glikemi atau gula darah hingga mendekati kadar gula darah yang mendekati normal. Bila upaya tersebut dapat dicapai penderita diabetes akan dapat dikondisikan menjadi lebih sehat (Rismayanti C., 2013)

Suntikan insulin adalah upaya terakhir bila semua upaya tidak berhasil dilakukan. Menurut Dr. Olly Renaldi dari RS Mitra Keluarga, paradigma baru penanganan penyakit diabetes adalah dengan pemberian insulin sejak awal sebelum terjadi kerusakan berat pada pancreas. Paradigma baru ini dikenal dengan istilah *Intensive Diabetes Management*. Dalam IDM, perubahan gaya hidup, pengaturan makan dan olah raga tetap dikedepankan. Pandangan bahwa insulin adalah upaya terakhir adalah pandangan yang kurang tepat, karena semakin lambat pemakaian insulin maka akan semakin parah terjadi kerusakan pankreas. Bila fungsi pankreas dinilai sudah mampu memproduksi insulin dengan baik maka suntikan insulin dari luar dapat dihentikan (www.mitrakeluarga.com).

HIDDEN NAÏVE BAYES

Klasifikasi jenis penyakit Diabetes Mellitus pada penelitian kali ini menggunakan algoritma Hidden Naïve Bayes yang merupakan pengembangan dari Naïve Bayes Classifier (Zhang H., Jiang L., Su J., 2005).

Naïve Bayes Classifier, dengan biaya komputasi yang minimal menghasilkan performa klasifikasi yang cukup baik, oleh karena itu Naïve Bayes dan pengembangannya sering digunakan sebagai algoritma untuk klasifikasi data. Pada penelitian ini akan diuji algoritma Hidden Naïve Bayes untuk klasifikasi data penyakit Diabetes Mellitus.

Naïve Bayes Classifier adalah bentuk paling sederhana dari Bayesian Network. Model strukturalnya adalah sebuah directed graph di mana node merepresentasikan atribut dan panah merepresentasikan dependency atributnya (Zhang H., Jiang L., Su J., 2005). Misalkan A_1, A_2, \dots, A_n adalah n atribut. Misalkan nilai A diwakili oleh a_1, a_2, \dots, a_n dan C adalah kelas datasetnya dan c adalah nilai C maka Bayesian Classifier dapat direpresentasikan dalam formula berikut ini :

$$c(E) = \arg \max_{c \in C} P(c)P(a_1, a_2, \dots, a_n|c). \quad (1)$$

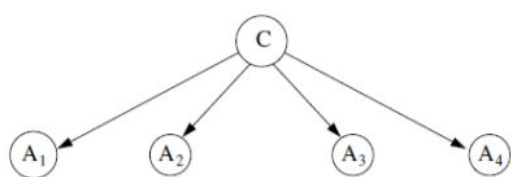
Bayesian Classifier mengasumsikan bahwa semua atribut independent terhadap kelas maka :

$$P(E|c) = P(a_1, a_2, \dots, a_n|c) = \prod_{i=1}^n P(a_i|c). \quad (2)$$

Sehingga hasil akhir klasifikasi Naïve Bayes kemudian dapat dirumuskan dalam formula sebagai berikut :

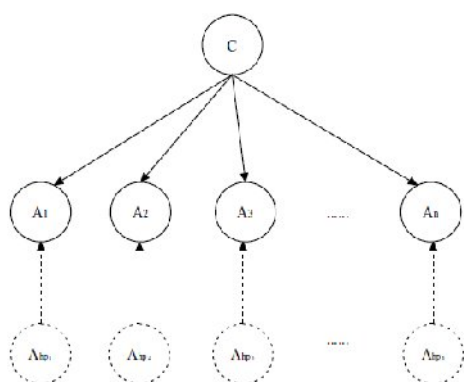
$$c(E) = \arg \max_{c \in C} P(c) \prod_{i=1}^n P(a_i|c). \quad (3)$$

ASUMSI NAÏVE BAYES yang merupakan bentuk paling sederhana dari Bayesian Networ (seperti tampak pada Gambar 2.1) bahwa atribut bersifat *conditional independence* sebenarnya kurang tepat.



Gambar 1 Struktur Naïve Bayes
Sumber : Hidden Naïve Bayes (Zhang H., Jiang L., Su J., 2005)

Hidden Naïve Bayes adalah algoritma yang merupakan pengembangan Naïve Bayes yang tidak mengabaikan independensi antar atributnya sehingga dibuat sebuah parent untuk tiap atribut yang bersifat hidden (Zhang H., Jiang L., Su J., 2005). Struktur sebuah Hidden Naïve Bayes dapat digambarkan seperti dalam Gambar 2 berikut ini :



Gambar 2 Struktur Hidden Naïve Bayes
Sumber : Hidden Naïve Bayes (Zhang H., Jiang L., Su J., 2005)

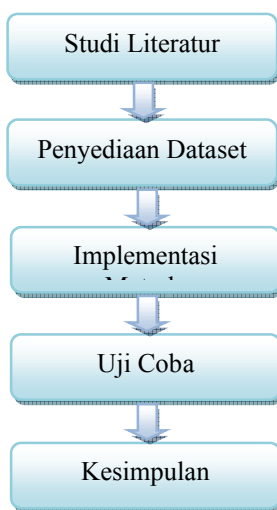
METODOLOGI PENELITIAN

Langkah-langkah atau tahap-tahap kegiatan penelitian dijelaskan dalam Blok Diagram seperti tampak pada Gambar 3. Pada tahap pertama dilakukan studi literature yang bertujuan untuk menjelaskan kajian pustaka dari teori-teori penunjang yang mendukung konstruksi penelitian. Dengan studi literature diharapkan dapat dijelaskan secara lebih baik berkaitan dengan teori-teori penunjang yang membahas

mengenai penyakit diabetes dan juga algoritma Hidden Naïve Bayes yang merupakan pengembangan dari Naïve Bayes Classifier. Juga dipelajari tentang beberapa parameter uji yang akan menjadi alat evaluasi kinerja algoritma yang digunakan dalam penelitian.

Pada tahap berikutnya dilakukan penyediaan dataset. Pada penelitian kali ini dataset yang digunakan adalah dataset Diabetes yang bisa diunduh secara gratis dari UCI Machine Learning Repository. Dataset Diabetes yang disediakan di UCI Machine Learning Repository memiliki 8 atribut yaitu preg, plas, pres, skin, insu, mass, pedi dan age serta 2 kelas yaitu tested positive sebanyak 500 dataset dan tested negative sebanyak 270 dataset.

Selanjutnya implementasi metode dan pengujian dilakukan dengan tool Weka versi 3.6.9 untuk melakukan skema klasifikasi berbasis Hidden Naïve Bayes. Hasilnya juga akan dibandingkan dengan beberapa algoritma lain yang serumpun dengan Naïve Bayes yaitu Bayes Net, Naïve Bayes sendiri dan AODE. Parameter uji yang akan digunakan antara lain adalah : TP Rate, FP Rate, Precision, Recall dan F Measure. Sedangkan dua macam scenario uji yang akan digunakan adalah training tanpa cross validasi dan training dengan cross validasi.



Gambar 3.1 Langkah-langkah Kegiatan Penelitian

DATASET

Dataset yang digunakan untuk pengujian dalam penelitian ini adalah dataset Diabetes yang diunduh dari UCI Machine Learning Repository. Owner dataset diabetes adalah National Institutes of Diabetes and Digestive and Kidney Diseases yang merupakan kontribusi dari Vincent Sigillito Research Center, RMI Group Leader Applied Physics Laboratory The John Hopkins University. Jumlah fitur dataset Diabetes adalah 8 fitur yaitu : preg, plas, pres, skin, insu, mass, pedi dan age. Jumlah kelasnya ada dua yaitu tested_positive dan tested_negative. Jumlah keseluruhan dataset yang digunakan untuk pengujian metode Hidden Naïve Bayes (HNB) adalah sebanyak 668 data di mana jumlah kelas tested_positive nya sebanyak 268 data dan jumlah kelas tested_negative nya sebanyak 500 data.

ANALISA DAN PEMBAHASAN

Hasil klasifikasi dataset diabetes dengan algoritma Hidden Naïve Bayes dengan scenario pengujian dengan menggunakan training set dan menggunakan 10 cross validasi ditunjukkan dalam confusion matrix yang disajikan dalam Tabel 4.1. berikut ini :

Tabel 1 Hasil Klasifikasi HNB dengan Training Set

Terklasifikasi sebagai	tested positive	tested_negative
tested_positive	268	0

tested_negative	0	500
-----------------	---	-----

Jumlah data kelas tested_positive dalam dataset adalah 268 dan terklasifikasi dengan HNB sebagai tested_positive sebanyak 268, sedangkan jumlah data kelas tested_negative dalam dataset adalah sejumlah 500 dan terklasifikasi sebagai tested_negative sebanyak 500 sehingga sehingga dapat dihitung nilai parameter uji berupa TP Rate, FP Rate, Precision, Recall dan F-Measure seperti ditampilkan dalam Tabel 4.2.

Tabel 2 Performance Klasifikasi HNB dengan Training Set

Kls/Par	TP Rat e	FP Rat e	Prec	Rec	F Meas
tested_positive	1	0	1	1	1
tested_negative	1	0	1	1	1
Rata-rata	1	0	1	1	1

Hasil tersebut lebih baik bila dibandingkan dengan klasifikasi dengan Naïve Bayes.

Tabel 3 Hasil Klasifikasi Naïve Bayes dengan Training Set

Terklasifikasi sebagai	tested positive	tested_negative
tested_positive	238	30
tested_negative	25	475

Dari convision matrix di atas, bila dihitung rata-rata kinerja algoritma Naïve Bayes dapat ditampilkan dalam Tabel 4.

Tabel 4 Performa Klasifikasi Naïve Bayes dengan Training Set

Kls/Par	TP Rate	FP Rate	Prec	Rec	F Meas
tested_positive	0.888	0.050	0.905	0.888	0.896
tested_negative	0.950	0.112	0.941	0.950	0.945
Rata-rata	0.919	0.081	0.923	0.919	0.921

KESIMPULAN

Dari hasil pengujian dapat disimpulkan beberapa hal sebagai berikut :

1. Algoritma Hidden Naïve Bayes adalah penyempurnaan Naïve Bayes Classifier
2. Algoritma Hidden Naïve Bayes dapat dipergunakan untuk klasifikasi data Diabetes
3. Hidden Naïve Bayes menunjukkan kinerja klasifikasi yang lebih baik dibandingkan dengan Naïve Bayes Classifier dalam pengujian dengan training set.

DAFTAR PUSTAKA

- [1] Yoon, K. H., et al., “*Epidemic Obesity and Type 2 Diabetes in Asia*”, Lancet 2006 Division of Endocrinology and Metabolism, College of Medicine, Catholic University of Korea.
- [2] Tuomilehto, J., et al., “Prevention of Type 2 Diabetes Mellitus by Changes in Lifestyle among Subjects with Impaired Glucose Tolerance”, *The New England Journal of Medicine* Vol. 344 Number 18 : 2001
- [3] Rismayanti C., “Terapi Insulin sebagai Alternatif Pengobatan Bagi Penderita Diabetes”, Repository Universitas Negeri Yogyakarta
- [4] www.mitrakeluarga.com diakses pada 11 Juli 2013
- [5] Zhang, H., Jiang, L., & Su, J. (2005, July). Hidden naive bayes. In *Proceedings of the National Conference on Artificial Intelligence* (Vol. 20, No. 2, p. 919). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.
- [6] Shaw J. E., Sicree R. A., Zimmet P. Z., “Global Estimates of The Prevalence of Diabetes”, *Diabetes Research and Clinical Practice* 87 (2010) 4-14
- [7] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1), 10-18.

#Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember

¹bagus.setya@unmuhjember.ac.id