KLASIFIKASI PENDERITA PENYAKIT DIABETES MELITUS MENGGUNAKAN ALGORITMA *ADABOOST*

JURNAL

FINGKI IBRAHIM
A01419036

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN ILMU KOMPUTER
UNIVERSITAS MUHAMMADIYAH GORONTALO
GORONTALO
2024

KLASIFIKASI PENDERITA PENYAKIT DIABETES MELITUS MENGGUNAKAN ALGORITMA *ADABOOST*

Fingki Ibrahim[1], Rubiyanto Maku[2], Wahyudin Hasyim[3]
[1,2,3] Program Studi Sistem Informasi, Universitas Muhammadiyah Gorontalo, *Jl. Prof. Dr. Mansoer Pateda, Kab.Gorontalo , India Email :*
fingkiibrahim9@gmail.com

ABSTRACT

FINGKI IBRAHIM, Diabetes Melitus Patient Classification Using the Adaboost Algorithm. Under the direction of WAHYUDIN HASYIM as supervisor 2 and RUBIYANTO MAKU, as supervisor 1,

Diabetes is a condition in which the body's blood sugar or glucose levels are higher than normal. Problems in the body's organs result from cells' inability to adequately absorb glucose, which leads to an accumulation of the substance in the blood. The use of the Adaboost algorithm in the categorization of diabetes is the goal of this study. To provide an accurate classification, this study undertook a number of steps, including data collection, attribute selection, model development, and classification outcomes. Based on the data used, the Adaboost algorithmm's accuracy rate is 96.87% according to the results.

Keywords: classification, data mining, diabetes mellitus, and Adaboost

ABSTRAK

FINGKI IBRAHIM, Klasifikasi Penderita Penyakit Diabetes Melitus Menggunakan Algoritma Adaboost. Dibimbing Rubiyanto Maku, M.Kom sebagai pembimbing 1 dan Wahyudin Hasyim, M.Kom sebagai pembimbing 2

Diabetes adalah penyakit dimana glukosa atau gula darah tubuh melebihi batas normal dalam tubuh. Akumulasi glukosa dalam darah disebabkan oleh sel-sel yang tidak dapat mengambilnya dengan benar dalam tubuh yang menyebabkan ganguan pada organ tubuh. Tujuan penelitian ini adalah Penerapan Algoritma Adaboost dalam klasifikasi penyakit diabetes. Pada penelitian ini, dilakukan beberapa tahap untuk menghasilkan klasifikasi yang akurat yaitu: pengumpulan data, pemilihan atribut, pembangunan model dan hasil klasifikasi. Hasil menunjukkan bahwa algoritma Adaboost memiliki tingkat akurasi mencapai 96.87%. pada data yang digunakan.

Kata Kunci : Diabetes melitus, Data Mining, Klasifikasi, Adaboost

1. Latar Belakang

Pada akhir tahun 2021, *internasional diabetes federation (IDF)* dalam Atlas edisi ke-10 mengkonfirmasi bahwa diabetes termasuk salah satu keadaan darurat kesehatan global yang tumbuh paling cepat di abad ke-21. Pada tahun 2021, lebih dari setengah miliar orang diseluruh dunia hidup dengan diabetes, atau 537 juta tepatnya, dan jumlah ini diperkirakan akan mencapai 643 juta pada tahun 2030 dan 783 juta pada tahun 2045. Ada juga juga populasi besar penderita diabetes, diperkirakan jumlah orang yang gula darahnya mulai naik atau berada di tahap pradiabetes, yaitu gangguan toleransi glukosa pada tahun 2021 adalah sekitar 541 jura orang. Diabetes pada populasi ini juga menyebabkan tingginya angka kematian terkait diabetes, diperkirakan lebih 6,7 juta pada orang dewasa berusia 20-79 tahun. (Dr. dr. Made Ratna Saraswati, SpPD-KEMD, n.d.)

Diabetes adalah penyakit dimana glukosa atau gula darah tubuh melebihi batas normal dalam tubuh. Akumulasi glukosa dalam darah disebabkan oleh sel-sel yang tidak dapat mengambilnya dengan benar dalam tubuh yang menyebabkan ganguan pada organ tubuh. Diabetes juga merupakan penyakit yang menyerang pada gangguan metabolism kronis. Pada penderita diabetes,pancreas tidak mampu memproduksi insulin sesuai dengan kebutuhan tubuh, termasuk memungkinkan glukosa untuk digunakan sebagai energy di tubulus. Menurut organisasi kesehatan dunia (WHO), hingga saat ini telah terjadi lebih dari 150 juta jiwa kematian mengidap penyakit tersebut, baik orang tersebut adalah rang tua, dewasa, remaja, pria atau wanita. Karena itu tidak heran jika penyakit diabetes ini menjadi perhatian dikarenakan jumlah penderita meningkat setiap tahunnya. Hal ini disebabkan fakta bahwa bentuk diabetes saat ini hanya dapat dideteksi bila terdapat komplikasi pada saluran kemih. Untuk itu, anda perlu mewaspadai kondisi diabetes anda sejak dini. (Novianti et al., 2022)

2. LANDASAN TEORI

1. Data

Data adalah kumpulan fakta Diperoleh dari pengukuran. Sesuatu mengambil keputusan yang tepat adalah hasil dari penarikan kesimpulan berdasrkan data atau fakta. Kita dapat mengatakan data itu fakta yang menggambarkan sesuatu peristiwa atau bentuk yang masih yang memerlukan pemrosesan lebih lanjut sehingga dapat menghasilkan informasi. (Putry et al., 2022)

2. Data Mining

Metode yang digunakan untuk menggali Informasi yang tidak ditemukan di tempat lain disebut dengan data mining. Sejak 1990,Istilah data mining sudah tidak asing lagi Karena itu perlu untuk mengolah data bermanfaat dan sangat diperlukan dilakukan oleh departemen yang berbeda, di kesehatan, sains, dan bahkan industri. Ilmu data mining adalah perpaduan ilmu dari Kecerdasan Buatan, Sains, Statistik danpenelitian database yang semakin luas.Menurut metode data mining pada artikel adalah proses pengikatan Ini termasuk makna, pola dan hubungan dengan memproses kumpulan data. Dalam data mining Ada 6 metode yang umum digunakan yaitu memprediksi atau meramalkan, penggambaran atau deskripsi, klasifikasi, estimasi, asosiasi dan *clustering*.(Kudus *et al*., 2020).

3. Klasifikasi

Kalsifikasi merupakan teknik pengolahan data yang membagi objek menjadi beberapa kelas sesuai dengan jumlah kelas yang diinginkan. Klasifikasi adalah teknik untuk menemukan model yang dapat memisahkan satu kelas data dari yang lain untuk menentukan objek mana yang termasuk dalam kategori tertentu dengan memeriksa perilaku dan sifat dari kelompok yang dientukan. Teknik ini

dapat mengklasifikasikan data baru menggunakan hasil untuk memberikan beberapa aturan.(Romli et al., 2020)

4. *Adaboost*

Adaptive Boosting (Adaboost) adalah salah satu varian dari algoritma boosting. Adaboost adalah jenis pembelajaran ensambel yang sering diterapkan untuk meningkatkan algoritma boosting. Boosting bisa dikombinasikan dengan beberapa algroritma *classifier* untuk meningkatkan kinerja klasifikasi. Itu membuat pengertian bahwa jika model berbeda dari yang lain, beberapa penambahan model akan membantu. Adaboost dan variasinya telah berhasil di sejumlah bidang(domain) karena landasan teoritisnya yang kuat, prediksi yang akurat, dan kesederhanaan yang besar. (Laila Qadrini, 2021)

Langkah-langkah dalam Algoritma *Adaboost* adalah sebagai berikut:

1. Kumpulan himpunan data N sampel dengan dua kelas berlabel y {0,1}.
2. Dalam himpunan data, bobot awal setiap sampel ditetapkan sama, yaitu:

$$W_i^1 = \frac{1}{N} \text{ …….. (2.1)}$$

Dengan
$W_i^1 =$ Bobot awal pada iterasi pertama untuk semua sampel, dan
$i = 1, 2, \ldots, N.$

3. Untuk tiap iterasi t =1, 2,…T dengan T adalah iterasi maksimum, lakukan proses berikut:
   a) Penjumlahan kesalahan pembobotan dari semua sampel dihitung dengan sebagai berikut:
   $error\ \sum_{i=1}^{N}(w_i^t \times error_i^t$ …….(2.2)

   b) Indeks kepercayaan dari learner $f(x)$ di hitung sebagai berikut:

$$a^t = \frac{1}{2}\ln\left(\frac{1-error^t}{erorr^t}\right). \text{ …… (2.3)}$$

Dalam hal ini, indeks kepercayaan dari model prediktif $f(x)$ bergantung pada kesalahan pembobotan.

c) Memperbarui bobot untuk semua training sampel asli:

$$w_i^{t+1} = w_i^t \begin{cases} exp\ (-a^t)\ untuk\ menurunkan\ bobot \\ exp\ (a^t)\ untuk\ menaikan\ bobot. \end{cases}$$

Bila sampel diklasifkasikan benar, maka bobotnya akan menurun, sedangkan jika sampel diklasifikasikan salah, maka bobotnya meningkat.

d) Selanjutnya, dilakukan normalisasi bobot sebagai berikut:

$$w_i^{*t} = \frac{w_i^t}{\sum_i^N w_i^t} \text{ …… (2.4)}$$

Sehingga,

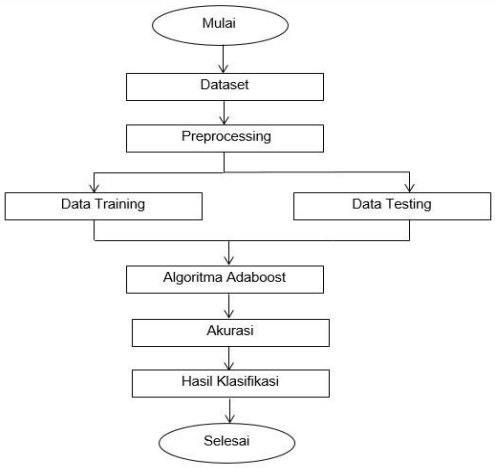
$$\sum_i^N w_i^{*\,t+1} = 1 \text{ …… (2.5)}$$

e) Jika error < 0,5, atau *t* < T, ulangi langkah 1-4; jika tidak maka proses berhenti.
f) Model prediksi f^t (x) dengan t = 1,2,…, T diperoleh setelah T iterasi. Peramalan akhir untuk kasus i ditentukan oleh peramalan model T yang digabungkan dengan menggunakan pendekatan voting, yaitu sebagai berikut:

$$\sum_{t=1}^{T} a^t\ f^t\ (x) \text{ …….. (2.6)}$$

3. METODOLOGI PENELITIAN

Adapun tahapan-tahapan yang akan digunakan dalam penelitian ini dapat dilihat pada gambar 1 berikut :



Tabel 1. Tahapan Penlitian

4. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Penelitian ini menggunakan data yang bersifat publik. Data tersebut diambil dari kumpulan data yang memprediksi risiko diabetes tahap awal, serta mencakup tanda dan gejala pada individu yang baru didiagnosis atau berisiko menjadi penderita diabetes. Data ini dikumpulkan menggunakan kueisoner langsung dari pasien rumah sakit diabetes Sylhet di Syhet, Bangladesh dan disetujui oleh dokter. Data ini mencakup jumlah gejala penderita diabetes atau yang akan menjadi penderita diabetes pada tahun 2021, dimana penderita yaitu laki-laki dan perempuan pada usia 20-65 tahun. Label menunjukan possitive dan negative.

Dataset yang digunakan dalam penelitian ini adalah dataset dari kaggle yang bisa ditemukan dihalaman web https://www.kaggle.com/datasets/ishandutta/early-stage-diabetes-risk-prediction-dataset?resource=download&select=diabetes_data_upload.csv

B. Prepocessing

Sebelum masuk pada tahap uji coba dengan *tool RapidMiner* serta perhitungan manual menggunakan algoritma *Adaboost* hal yang perlu dilakukan adalah merubah dataset yang awalnya berupa nominal menjadi *numeric*. Tahapan *preprocessing* adalah merubah data nominal menjadi *numeric* sehingga data dapat diolah dan menghasilkan akurasi yang diharapkan..

Tabel 2. Data hasil *Preprocessing*

No	Age	A.	B.	C.	D.	E.	F.	G.	H.	I.	J.	K.	L.	M.	N.	O.	P.
1.	40	1	0	1	0	0	0	0	0	1	0	1	0	1	1	1	Positive
2.	58	1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	Positive
3.	41	1	1	0	0	0	1	0	0	1	0	1	0	1	0	0	Positive
4.	45	1	0	0	1	0	1	1	0	1	0	1	0	0	0	0	Positive
5.	60	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	Positive
6.	55	1	1	1	0	0	1	0	1	1	0	1	0	1	1	1	Positive
7.	57	1	1	1	0	0	1	1	0	0	0	1	1	0	0	0	Positive
8.	66	1	1	1	1	0	0	0	1	1	1	0	1	1	0	0	Positive
9.	67	1	1	1	0	0	1	1	0	1	1	0	1	1	0	1	Positive
10.	40	1	0	1	1	0	1	0	1	1	1	0	0	0	1	0	Positive
...		...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
520	42	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	Negative

Ada beberapa atribut yang tidak dapat diubah yaitu umur karena umur tidak dapat diubah di *RapidMiner*. Dari data yang ada, yang dapat diubah kita singkat jadi A, B, sampai O yaitu Usia, *Poliuria* (sering buang air kecil) , *Polydipsia* (sering haus), penurunan berat badan mendadak, kelemahan, *polifigia* (Makan berlebihan), sariawan kelamin, visual kabur, gatal, *iritabilitas*, penyembuhan tertunda, *paresis parsial* (kelumpuhan), kekakuan otot, *alopesia* (rambut rontok), obesitas. Kelas tidak dapat diubah karena kelas adalah label atau atribut yang akan di jadikan target. Untuk *age* (A) berisi jawaban laki-laki kita beri 1 untuk perempuan kita beri 0. Yang berisi jawaban *Yes* yaitu 1 dan *No* 0.

1. Pengujian Akurasi Algoritma *Adaboost*

X-VALIDASI	ACCURACY
2	89.90%
3	93.75%
4	95.43%
5	96.40%
6	96.15%
7	95.90%
8	96.39%
9	94.71%
10	96.87%

Tabel 3. Pengujian Accuracy Adaboost

Berdasarkan table 2. Terlihat bahwa percobaan dengan 10 variabel periode dengan menggunakan *Cross Validation* dan mengganti nilai *input number of validation* = 2-10. Dari hasil percobaan yang dilakukan dengan menggunakan algoritma *Adaboost*, maka diperoleh nilai tertinggi pada *x-validation* = 10 *accuracy* 96.87%+/- 2.80%%(mikro: 96.88%) *class precision* 98.41% dan *class recall* 96.48%.

C. Hasil Klasifikasi

Pada tahapan ini peneliti menggunakan data training sebanyak 416 data, dan testing 104 data. Dengan menggunakan X-Validation = 10 maka hasil klasifikasi algoritma Adaboost menggunakan C4.5 untuk mengklasifikasikan penyakit diabetes melitus dapat dilihat pada gambar berikut ini.

	true Positive	true Negative	class precision
pred. Positive	557	13	97.72%
pred. Negative	3	467	99.36%
class recall	99.46%	97.29%	

Gambar 1. Hasil Klasifikasi Adaboost

Untuk hasil klasifikasi Adaboost menggunakan data *testing* 20% atau 104 data. Dan menggunakan operator *Cross Validation* dengan *Number Of Fold* 10 kali pengujian. Jadi total data *testing* menjadi 1040 data karena dikali 10.

5. KESIMPULAN

Berdasarkan pada hasil penelitian yang telah dilakukan maka dapat disimpulkan sebagai berikut :

1. Algoritma Adaboost yang digunakan dalam melakukan klasifikasi penderita penyakit diabetes mellitus menghasilkan akurasi sebesar 96.87%. nilai akurasi ini lebih tinggi dibandingkan dengan nilai akurasi dari sebagian besar penelitian-penelitian sebelumnya.
2. Dari Algoritma Adaboost menggunakan algoritma C4.5 berhasil divalidasi dan di implementasikan dapat mengklasifikasikan penyakit diabetes melitus. Berdasarkan data *training* dan *testing* langkah berikutnya melakukan klasifikasi sehingga menghasilkan kesimpulan bahwa jumlah data

training mempengaruhi pada presentase keakurasian.

3. Penerapan algoritma Adaboost untuk mengklasifikasi data penyakit diabetes melitus mendapatkan hasil akurasi yang memuaskan dengan jumlah *record* data yang banyak.

DAFTAR PUSTAKA

Abdurrahman, G. (2022). Jurnal Sistem dan Teknologi Informasi Klasifikasi Penyakit Diabetes Melitus Menggunakan Adaboost Classifier. 7(1).

Dr. dr. Made Ratna Saraswati, SpPD-KEMD, F.-R. P. dr. I. G. N. G. N. (2022). *Diabetes Melitus Adalah Masalah Kita*. Kementrian Kesehatan Direktorat Jendral Pelayanan Kesehatan. https://yankes.kemkes.go.id/view_artikel/1131/diabetes-melitus-adalah-masalah-kita

Novianti, N., Zarlis, M., & Sihombing, P. (2022). Penerapan Algoritma Adaboost Untuk Peningkatan Kinerja Klasifikasi Data Mining Pada Imbalance Dataset Diabetes. 6(April), 1200 12 06. https://doi.org/10.30865/mib.v6i2.4017

Putry, N. M., Sari, B. N., Kom, M., Informatika, T., & Karawang, U. S. (2022). KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELITUS. 10(1).

Kudus, U. M., Ganesha, J., & Kudus, P. (2020). Fida Maisa Hana. Jurnal Sistem Komputer Dan Kecerdasan Buatan (Siskom-KB), 4(1), 32–39.

Achmad Ridwan. (2020). *Aplikasi Monitoring Keamanan Ruangan Menggunakan IP Camera Berbasis Android Sri Apriyani , Ridho Taufiq Subagio , Wanda Ilham Pengukuran Panjang-Berat Ikan dan Sayuran secara Otomatis pada Budikdamber (Budi Daya Ikan dalam Ember) Menggunakan Visi Komput*. *IV*(1).

Amien, I. L. F., Astuti, W., & Lhaksamana, K. M. (2023). Perbandingan Metode Naïve Bayes dan KNN (K-Nearest Neighbor) dalam Klasifikasi Penyakit Diabetes. *E-Proceeding of Engineering*, *10*(2), 1911–1920.

ayo sehat Kementerian Kesehatan Republik Indonesia. (2023). *diabetes melitus tipe 2 adalah penyakit kronis yang ditandai dengan peningkatan kadar gula darah dalam tubuh*. Kementerian Kesehatan Republik Indonesia. https://ayosehat.kemkes.go.id/topik-penyakit/diabetes--penyakit-ginjal/diabetes-melitus-tipe-2

Hidayat, T., Anelia, S. S., Pratiwi, R. I., Salsabila, N., & Prasvita, D. S. (2020). Perbandingan Akurasi Klasifikasi Penyakit Diabetes Menggunakan Algoritma Adaboost-Random Forest Dan Adaboost-Decision Tree Dengan Imputasi Median Dan Knn. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, *April*, 616–623.

IslamEmail, M. F., Ferdousi, R., Rahman, S., & Bushra, H. Y. (2020). *Kumpulan Data Prediksi Risiko Diabetes Tahap Awal Data Tanda dan Gejala Penderita Diabetes Baru atau yang Akan Menjadi Penderita Diabetes*. Kaggle.Com. https://www.kaggle.com/datasets/ishandutta/early-stage-diabetes-risk-prediction dataset?resource=download&select=diabetes_data_upload.csv

Karyawan, D., Algoritma, M., Romli, I., & Zy, A. T. (2020). *Penentuan Jadwal Overtime Dengan Klasifikasi*. *4*(September), 694–702.

P2PTM Kemenkes RI. (2019). *Apa saja Gejala Penyakit Diabetes Melitus (DM)?*

Penyakit Tidak Menular Indonesia. https://p2ptm.kemkes.go.id/infographic-p2ptm/penyakit-diabetes-melitus/apa-saja-gejala-penyakit-diabetes-melitus-dm

Rahayu, D. S., Afifah, J., & Intan, S. (2023). *Classification of Diabetes Mellitus Using C4 . 5 Algorithm , Support Vector Machine (SVM) and Linear Regression Klasifikasi Penyakit Diabetes Melitus Menggunakan Algoritma C4 . 5 , Support Vector Machine (SVM) dan Regresi Linear*. 56–63.

Rais, A. N., Rahmawati, E., & Amir, R. F. (2021). *Prediksi Pima Indians Diabetes Database Dengan Ensemble Adaboost Dan Bagging*. *9*(2), 36–42.

Subarkah, P. (2020). *Penerapan Algoritma Klasifikasi Classification And Regression Trees (CART) untuk Diagnosis Penyakit Diabetes Retinopathy*. *19*(2), 294–301.

upk kemenkes. (2023). *Diabetes Melitus Tipe 2*. Unit Pelayanan Kesehatan Kemenkes. https://upk.kemkes.go.id/new/home

Laila Qadrini, Andi Sepperwali, dan Asra Aina. 2021. "Decision Tree Dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial." Jurnal Inovasi Penelitian 2(7):1959–66.