

PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI PENYAKIT DIABETES MELLITUS

Achmad Ridwan

Sistem Informasi, Universitas Muhammadiyah Kudus
Purwosari, Kudus, Indonesia

Email : [Achmad Ridwan@umkudus.ac.id](mailto: Achmad Ridwan@umkudus.ac.id)

Abstrak

Diabetes Mellitus merupakan penyakit turunan disebabkan oleh kadar gula tinggi dalam darah. Gula darah disimpan dan digunakan untuk energi yang dipindahkan ke sel manusia yang berasal dari darah oleh hormon insulin . ketika pasien mengalami Diabetes, insulin pada manusia tidak biasa dihasilkan secara seimbang dan tubuh tidak dapat menggunakan insulin tersebut secara tepat untuk kebutuhan badan. Diabetes Mellitus sebagai penyakit yang merupakan termasuk deretan penyakit penyebab kematian terbesar didunia. Diabetes Mellitus dapat diklasifikasikan berdasarkan kemungkinan mengalami diabetes dari atribut gejala yang muncul diawal fasenya. Penyakit ini bisa dideteksi kemunculannya karena banyak atribut gejala yang terdeteksi di awal fasenya. Dataset ketika analisis penelitian ini dilakukan merupakan data dari dataset UCI Machine Learning yaitu Early Stage Diabetes Risk pada tahun 2020 dan terdiri dari 17 atribut. Analisis penelitian ini antara lain Data Preprocessing, Modeling , dan Evaluasi. Model klasifikasi penelitian memakai Algoritma C4.5. Hasil Klasifikasi pada riset ini didapatkan akurasi sebesar 94.23% dan nilai AUCnya yaitu 0,952

Kata Kunci : Data Mining, Klasifikasi, C4.5, Diabetes Mellitus

Abstract

Diabetes Mellitus is a hereditary disease caused by high blood sugar levels. Blood sugar is stored and used for energy which is transferred to human cells which comes from the blood by the hormone insulin. when a patient has diabetes, insulin in humans is not normally produced in a balanced manner and the body cannot use the insulin properly for the body's needs. Diabetes Mellitus as a disease which is one of the largest causes of death in the world. Diabetes Mellitus can be classified based on the likelihood of developing diabetes from the attributes of symptoms that appear at the beginning of its phase. The emergence of this disease can be detected because many attribute symptoms are detected at the beginning of its phase. The data used in the analysis of this research is data from the UCI Machine Learning dataset, namely Early Stage Diabetes Risk in 2020 and consists of 17 attributes. The analysis of this research includes Data Preprocessing, Modeling, and Evaluation. The research classification model uses the C4.5 Algorithm. This classification results in the research obtained an accuracy of 94.23% and an AUC value of 0.952

Keywords : Data Mining, Classification, C4.5, Diabetes Mellitus

I. PENDAHULUAN

Insulin merupakan hormon yang diciptakan oleh pankreas, yang mempunyai fungsi seperti penyeimbang glukosa dari makanan yang dimakan kemudian mengalir ke sel-sel mulai dari darah didalam tubuh manusia yang menghasilkan energi. Diabetes merupakan penyakit ketika insulin tidak mampu diproduksi oleh pankreas atau tubuh manusia tidak dapat menggunakan insulin yang dibuat oleh Pankreas.

Makanan yang mengandung karbohidrat diproses menjadi glukosa dalam darah. glukosa dibantu insulin masuk ke dalam sel. Diwaktu tubuh tidak mampu memproduksi

insulin dan bahkan tidak mampu menggunakan dengan benar menyebabkan meningkatnya kadar glukosa dalam darah. Sehingga pada waktu jangka panjang, kadar glukosa tinggi dalam darah dapat merusak organ pada tubuh dan kegagalan fungsi organ dan jaringan. Peneliti membagi diabetes menjadi diabetes tipe satu, tipe dua, dan diabetes gestasional [1]. Diabet gestasional adalah jenis diabetes yang mungkin terjadi pada wanita hamil karena perubahan hormonal. Gejala yang umum diabetes adalah poliuria, polidipsia, polifagia, penurunan berat badan mendadak (biasanya tipe 1), kelemahan, obesitas (biasanya tipe 2),

alopecia, sariawan, gatal, penglihatan kabur, penyembuhan tertunda, iritabilitas, genital, paresis, parsial, otot kaku, dll. [2].

Di modern ini, teknologi dapat membantu untuk mendeteksi penyakit dengan akurat dan efisien waktu. Penggalian data adalah salah satu fungsi bidang ilmu komputer yang digunakan untuk prediksi atau estimasi. Ini merupakan proses menyaring data besar menjadi data ekstrak yang di kumpulkan dan melalui analisis data [3]. Semua proses tersebut dapat memprediksi suatu penyakit dengan pendekatan data mining. Data mining memerlukan gejala yang digabungkan dengan data klinis. Semua Gejala merupakan faktor yang amat penting untuk pasien baru dan prediksi tahap awal dengan data gejala. Penelitian ini juga membutuhkan data klinis untuk menganalisisnya

Salah satu pengklasifikasi data mining adalah algoritma C4.5. Metode algoritma C4.5 digunakan untuk mengklasifikasikan penyakit Diabetes Mellitus

Research ini menganalisis Klasifikasi Algoritma C4.5 dalam mengklasifikasi gejala awal Diabetes Mellitus sehingga mendapatkan akurasi yang di dapat dari hasil proses evaluasi.

A. Data Mining

Data Mining merupakan sebuah ilmu yang mempelajari proses kerja penggalian data dimana atributnya saling berkaitan agar menemukan pengetahuan atau menemukan pola data dari suatu data yang besar. Proses Penggalian data menuju pengetahuan dapat disimpulkan mempunyai 3 inti yaitu :

Data adalah fakta yang terekam dan belum membawa arti

Informasi adalah rangkuman fungsi penjelasan dan statistik dari data

Pengetahuan ini merupakan hasil dari penggalian dari data.

Nama yang sama untuk data mining adalah Knowledge Discovery in Data, Big data atau Business Intelijen, Knowledge Extraction dan Information Harvesting atau Pattern Analisis [4].

Inti dari data mining yaitu himpunan data yang telah terhimpun banyak sekali kemudian dengan metode data mining diproses dengan

rumus/algoritma data-data tersebut diambil atau diekstraksi untuk menjadi di sebuah pengetahuan. Pengetahuan Itulah bisa untuk diambil untuk sebuah tujuan tertentu. Pengaturan tersebut sangat berguna untuk berbagai keperluan .

B. Algoritma C 4.5

Algoritma C4.5 adalah salah satu algoritma untuk membuat decision tree berdasarkan data training yang sudah ada. Algoritma C4.5 adalah perkembangan ID3. perkembangan yang terjadi pada C4.5 adalah dapat mengatasi kerusakan data contohnya missing value, continue data, dan pruning data. Algoritma C4.5 merupakan salah satu algoritma top 10 yang terkenal [5] tetapi algoritma C4.5 belum bisa menangani ketidak seimbangan kelas [6].

Pada Pohon keputusan adalah metode penggalian data untuk prediksi dan klasifikasi yang paling kuat dan terbaik. Klasifikasi algoritma pohon keputusan dapat mengubah fakta yang ada dan besar menjadi algoritma pohon keputusan yang menggambarkan rule atau aturan. rule dapat mudah dipahami memakai bahasa alami. Mereka dapat digambarkan melalui bentuk bahasa basis data seperti SQL (Bahasa query) dalam mencari record pada kategori tersebut. Algoritma Pohon keputusan berfungsi juga untuk mendalami data seperti mencari hubungan yang belum terlihat diantara jumlah semua variabel masukan dengan variabel target dari dataset.

Algoritma pohon keputusan kolaborasi antara pendalaman data dan data modeling, pohon keputusan selalu bagus sebagai permulaan dalam proses data modeling sehingga ketika dijadikan model paling akhir dari semua teknik yang ada. Struktur pohon keputusan merupakan struktur yang sangat dapat dipakai dalam membagi sejumlah data yang sangat besar menjadi bagian kumpulan himpunan record yang kecil dengan mengujikan serangkaian aturan dari keputusan. Masing-masing serangkaian pembagian, bagian himpunan output menjadi sangat mirip satu dengan yang lainnya [7].

Model pada pohon keputusan terdiri dari kumpulan aturan yang berfungsi membagi sejumlah kumpulan yang beragam menjadi sedikit, lebih beragam dengan fokus pada

variabel outputnya. Pohon keputusan dibangun dengan teliti, lebih manual atau dapat berkembang dengan otomatis dan menggabungkan juga menerapkan beberapa model pohon keputusan untuk memodelkan kumpulan data yang masih belum terklasifikasi.

Semua Variabel output dikelompokkan dengan jelas dan sistem pohon keputusan lebih tertuju pada semua hitungan probability dari tiap tupel terhadap kategori yang ada atau berfungsi mengklasifikasi tupel dengan mengumpulkan di satu kelas. Pohon keputusan dapat berfungsi untuk memperkirakan nilai bersifat continue walaupun masih ada teknik yang sangat sesuai untuk kasus ini.

ID3, CART, dan C4.5 adalah contoh algoritma yang dapat dipakai untuk membentuk pohon keputusan. [8].

Bentuk tabel Data dalam pohon keputusan dinyatakan dalam bentuk tupel dan atribut. Atribut menyatakan sebuah parameter yang bertujuan agar terbentuk kriteria yang berfungsi pembentukan pohon. Contohnya Dalam menentukan main tenis, atribut yang dapat diperhatikan adalah cuaca, temperature, dan angin. Dari banyak Atribut yang merupakan atribut dapat dinyatakan sebagai data solusi dari data yang dinmakan target atribut . Atribut mempunyai nilai yang disebut instance. Contohnya atribut cuaca punya instance dapat berupa hujan, cerah, dan berawan, [7].

Jalan kerja pohon keputusan adalah merubah bentuk data berupa tabel menjadi model berupa pohon, merubah model berupa pohon menjadi inti rule, dan menjadikan sederhana rule tersebut [9]. Berikut ini dasar dari C4.5 :

Input berupa Data training dengan label training beserta atribut dengan Membuat simpul inti akar pohon yang akan dibuat pada algoritma ini:

1. Beri tanda (-), Jika semua sampel negatif, berhentilah dengan sebuah pohon dengan satu node simpul akar
2. Beri tanda (+), Jika semua sampel positif, berhentilah dengan sebuah pohon dengan satu node simpul akar
3. Jika atribut tidak ada isinya,

berhentilah dengan sebuah pohon dengan sebuah simpul akar, dengan menulis label yang sesuai nilai terbanyak yang berada di label training

4. Untuk selanjutnya,yang lain dimulai
 - a. A sebagai atribut yang hasil terbaik pengklasifikasinya (sesuai Gain rasio)
 - b. Atribut sebagai keputusan berfungsi sebagai simpul node akar atribut A
 - c. Setiap untuk nilai v_i , kemungkinan untuk A

- 1) Tambahkan cabang di bawah akar yang berhubungan dengan $A = v_i$

- 2) Tentukan sampel S_{v_i} sebagai subset dari sampel yang mempunyai nilai v_i untuk atribut A

- 3) Jika data sample S_{v_i} kosong maka

- i. Pada di bawah cabang ditambah simpul note daun dengan label = nilai terbanyak yang ada di label data training

- ii.Data lain tambahkan cabang yang baru di bagian bawah cabang yang Ada

- d. selanjutnya Berhenti berakhir

Mengubah pohon yang ada di rule yang sudah terbentuk. Jika Jumlah path sama dengan jumlah pada rule kemungkinan dapat dibuat dari root hingga leaf node.

Proses Tree Pruning dilaksanakan untuk menyederhanakan pohon sampai akurasi bertambah. Tree Pruning ada dua, yaitu:

Pre-pruning, merupakan menghentikan pembuatan suatu subpohon di awal yaitu memutuskan untuk tidak dekat dengan memecah data training. Saat berhenti, node berubah jadi leaf . Pada Node yang akhir ini menjadi kelas yang sering muncul antara subset data sampel.

Post-pruning, adalah setelah tree selesai dibangun maka menyederhanakan pohon dengan cara membuang beberapa cabang subtree. Pada Node yang jarang dipotong akan menjadi leaf dengan kelas frekwensinya sering muncul.

Cara hitung nilai gain, entropy, dan split dari tiap atribut data training akan menghasilkan sebuah gain ratio. Cara menghitung gain, entropy, dan split dan gain ratio ada pada rumus berikut ini [10] :

$$Entropy(S) = \sum - p_i * \log_2 p_i \quad (1)$$

kasus yang sama dengan hasil diagnosis negatif.

Confusion matrix berfungsi untuk membuat penilaian kerja model klasifikasi yang mempunyai jumlah objek yang diramal dengan benar dan salah [13]. Akurasi kelas minoritas dapat menggunakan metrik recall. Rumus dapat digunakan dipengujian ini adalah:

Akurasi adalah perbandingan jumlah prediksi yang benar. Semua ditentukan dengan mengimplementasikan rumus akurasi berikut:

$$Accuracy = \frac{a+b}{a+b+c+d} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Sensitivity disebut juga dengan recall. Jika sensitivity 100% sama artinya dengan pengklasifikasian menganggap kasus yang diamati positif. Sebagai contoh semua orang memiliki kanker ganas diterka orang tersebut sakit.

$$Recall = \frac{\text{number of True Positive}}{\text{number of True Positive} + \text{number of False Negative}} \quad (6)$$

Presisi adalah tingkat positif salah adalah perbandingan nilai positif yang salah diklasifikasikan pada kasus negatif, yang perhitungannya menggunakan Rumus :

$$Precision = \frac{\text{number of true positive}}{\text{number of true positive} + \text{number of false positive}} \quad (7)$$

2) Kurva ROC

Pada Kurva ROC berfungsi untuk memperoleh nilai hasil sebuah prediksi, Kurva ROC adalah teknik untuk menarasikan, manajemen, dan memilih proses Klasifikasi berdasarkan kinerja sistem [8]. Kurva ROC merupakan perangkat dua dimensi di dunia data mining yang digunakan untuk memperoleh nilai kinerja klasifikasi yang terdiri dua class keputusan, setiap objek diplot ke salah satu elemen dari himpunan berpasangan, positif dan negatif. Pada kurva ROC, nilai True Positif rate dipasang pada sumbu Y dan nilai False Positif rate dipasang pada sumbu X. Menurut Gorunescu dalam metode klasifikasi data mining, Nilai AUC dapat klasifikasikan menjadi beberapa hasil :

1. 0,90 sampai 1,00 = Hasil Sangat Baik
2. 0,80 sampai 0,90 = Hasil Baik

3. 0,70 sampai 0,80 = Hasil Sedang
4. 0,60 sampai 0,70 = Hasil Buruk
5. 0,50 sampai 0,60 = Hasil Kegagalan

II. METODE PENELITIAN

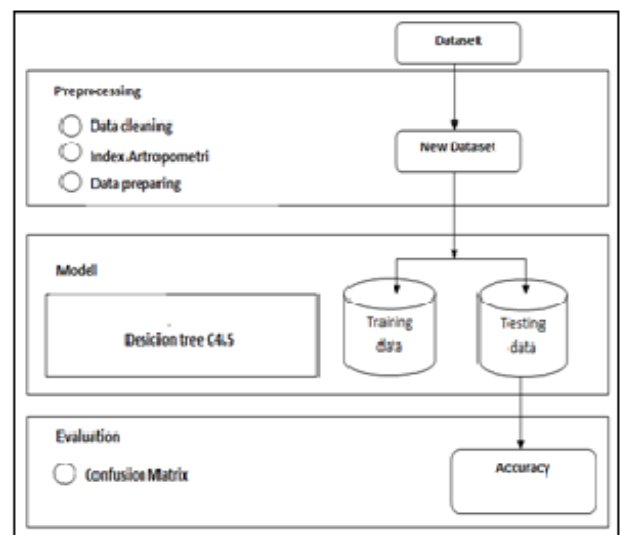
Metode dalam Research yaitu penelitian pendekatan kualitatif dan penelitian pendekatan kuantitatif. Dalam research ini metode pendekatan yang digunakan yaitu metode penelitian pendekatan kuantitatif. Tujuan dari penelitian ini adalah melakukan klasifikasi dan evaluasi model algoritma C4.5 untuk mengetahui akurasi klasifikasi algoritma C4.5 dalam mengklasifikasikan pasien yang terkena Diabetes Mellitus.

A. Sumber Data

Sumber dataset yang dipakai pada research ini adalah dataset dari dataset UCI Repository dataset. Dataset yang digunakan adalah Early stage diabetes risk prediction dataset tahun 2020. File tersebut bernama diabetes_data_upload.csv. Variabel yang digunakan pada research ini adalah sebanyak 17 variabel dengan jumlah data sebanyak 520. Ini termasuk data tentang orang-orang termasuk gejala yang dapat menyebabkan diabetes. Kumpulan data ini dibuat dari kuesioner langsung kepada orang-orang yang baru saja menjadi penderita diabetes, atau yang masih nondiabetes tetapi memiliki sedikit atau lebih gejala.

B. Kerangka Pikiran

Kerangka penelitian dari research bisa dilihat pada gambar 2.1 dibawah ini :



Gambar 2.1 Kerangka Pikiran

Adapun kerangka pikiran pada Gambar 2.1 dapat dinarasikan sebagai berikut :

1. Preprocessing

Preprocessing data dilakukan dengan cara menangani nilai yang hilang mengikuti teknik mengabaikan tupel dengan nilai yang tidak lengkap. Setelah praproses, total 500 instans telah tersisa. Diantaranya, 314 adalah nilai positif dan 186 adalah nilai negatif. Deskripsi detail dari dataset dan atributnya ditunjukkan pada Tabel 4.1. Dua variabel kelas digunakan untuk mengetahui apakah pasien memiliki risiko diabetes (positif) atau tidak (negatif).

2. Model

Software RapidMiner diterapkan pada Dataset yang baru untuk di Training maupun di Testing pada Algoritma C4.5, serta untuk menganalisis hasilnya antara lain Kesalahan klasifikasi (Classification Error), nilai akurasi Probabilitas maksimal tiap kelas, Recall dan Presisinya

3. Evaluasi

Dataset Diabetes Mellitus diuji/dievaluasi dengan Confusion Matrix serta diukur tingkat akurasi.

III. HASIL DAN PEMBAHASAN

Variabel data penelitian Dataset Diabetes Mellitus pada penelitian ini ditampilkan di Tabel 3.1 yakni sebagai berikut.

Tabel 3.1. Deskripsi Variable Dataset

No.	Atribut	Value
1	Age	1.20 – 35, 2.36 – 45, 3.46 – 55, 4.56 – 65, 6. lebih dari 65
2	Sex	1. Female, 2. Male
3	Polyuria	1. No , 2. Yes.
4	Polydipsia	1. No , 2. Yes.
5	Sudden weight loss	1. Nov, 2 Yes.
6	Weakness	1. No , 2. Yes.
7	Polyphagia	1. No , 2. Yes.
8	Genital thrush	1. No , 2. Yes.
9	Visual blurring	1. No , 2. Yes.
10	Itching	1. No , 2. Yes.
11	Irritability	1. No , 2. Yes.
12	Delayed healing	1. No , 2. Yes.
13	Partial paresis	1. No , 2. Yes.

14	Muscle stiffness	1. No , 2. Yes.
15	Alopecia	1. No , 2. Yes.
16	Obesity	1. No , 2. Yes.
17	Class	1. Positive, 2. Negative.

Dari Tabel 3.1 Ada 16 variabel dataset gejala dan 1 variabel class penentu klasifikasi

A. Preprocessing

Dari dataset yang akan di analisis dijadikan sebagai Data Training dan Data Testing yang ada di klasifikasikan oleh algoritma C4.5

Adapun contoh data untuk training dan data untuk testing di lihat pada tabel dibawah ini :

Tabel 3.2 Data Training dan data Testing

No.	Atribut	Value	
		1	2
1	Age	40	58
2	Sex	male	female
3	Polyuria	no	no
4	Polydipsia	yes	yes
5	Sudden weight loss	yes	yes
6	Weakness	yes	yes
7	Polyphagia	no	no
8	Genital thrush	no	no
9	Visual blurring	yes	yes
10	Itching	yes	yes
11	Irritability	no	no
12	Delayed healing	no	no
13	Partial paresis	yes	yes
14	Muscle stiffness	yes	yes
15	Alopecia	no	no
16	Obesity	no	no
17	Class	negatif	positif

Setelah praproses, total 500 data yang telah tersisa. Diantaranya, 314 adalah Class nilai positif dan 186 adalah Class nilai negatif.

B. Model

Setelah Preprocessing, data Training dan Data Testing akan diproses klasifikasinya menggunakan aplikasi Rapidminer Adapun hasil dari Confusion Matrix.nya

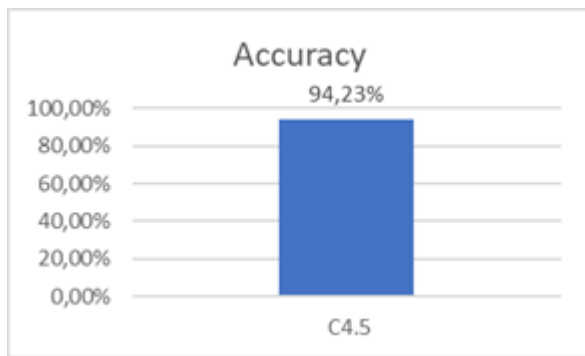
Tabel 3.3 Tabel Hasil Class Recall dan Precision

	Nilai true Positive	Nilai true Negative	Nilai class precision
Prediksi Nilai Positive	306	16	95.03%
Prediksi Nilai Negative	14	184	92.93%
class recall	95.62%	92.00%	

Dari Tabel 3.3 didapatkan Class Precision = 92.93%, dan Class Recall: 92.00%

C. Evaluasi

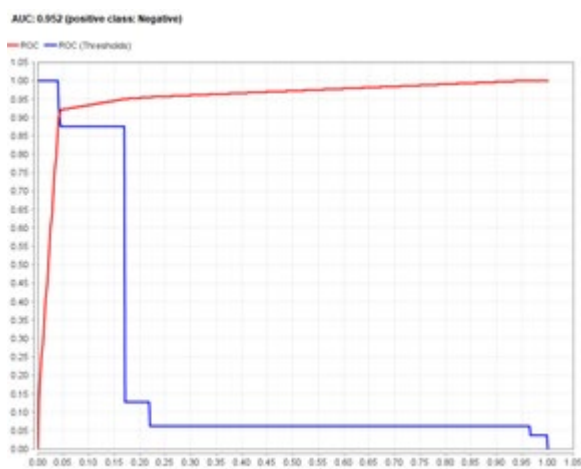
Hasil dari evaluasi pengklasifikasian dengan C4.5 menghasilkan :



Gambar 3.1 Akurasi Algoritma C4.5

Dari gambar diatas didapatkan akurasi untuk mengklasifikasikan Dataset Diabetes yaitu 94,23%

Adapun hasil dari pengujian ini juga menghasilkan gambar dibawah ini :



Gambar 3.2 Kurva Roc Model C4.5

Kurva ROC (Reciver Operating Characteristic) diatas menunjukkan algoritma C4.5memiliki nilai AUC sebesar 0.952 yang artinya Excellent Classification (Sangat Bagus) Dalam hasil penelitian, menunjukkan

algoritma C4. 5memberikan akurasi yang bagus untuk prediksi penyakit diabetes pada UCI dataset Machine Learning Repository.

IV. KESIMPULAN

Ada 16 atribut yang mempengaruhi Klasifikasi dataset Early stage diabetes risk prediction yaitu Gender, Age, Polyuria, Sudden Weight Loss, Polydipsia, Polyphagia, Weakness, Visual Blurring, Genital Thrush, Irritability, Itching, Delayed Healing, Muscle Stiffness, Partial Paresis, Alopecia, Class dan Obesity,.

Penelitian ini menggunakan Algoritma C4.5 untuk pengklasifikasian Dataset Early Stage Diabetes Risk Prediction. Dari 520 data awal setelah dilakukan pembersihan atau cleaning data yang belum lengkap dan data di proses seleksi disesuaikan yang dibutuhkan didalam analisis Dataset Stage Diabetes maka menghasilkan jumlah 500 data bersih.

Dari hasil proses Klasifikasi permodelan algoritma C4.5 terhadap Dataset Stage Diabetes menghasilkan Class Precision = 92.93%, dan Class Recall: 92.00%

Hasil evaluasi klasifikasi ini menghasilkan akurasi untuk mengklasifikasikan Dataset Diabetes yaitu 94,23% ini artinya Klasifikasi permodelan algoritma C4.5terhadap Dataset Stage Diabetes sudah bagus akurasinya, tetapi perlu peningkatan akurasi dengan Assemble atau dengan cara yang lain.

Kurva ROC (Reciver Operating Characteristic) Klasifikasi permodelan algoritma C4.5terhadap Dataset Stage Diabetes menunjukkan algoritma C4.5memiliki nilai AUC sebesar 0.952 yang artinya Excellent Classification (Sangat Bagus)

DAFTAR PUSTAKA

N. D. S. Report, “National Diabetes Statistics Report, 2020,” *Natl. Diabetes Stat. Rep.*, 2020.
 N. D. S. Report, “National Diabetes Statistics Report 2020 Estimates of Diabetes and Its Burden in the United States,” *Natl. Diabetes Stat. Rep.*, 2020.

- F. Gorunescu, *Data Mining*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
- J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- X. Wu and V. Kumar, *The Top Ten Algorithms in Data Mining*. Taylor & Francis Group, 2009.
- M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*. 2012, doi: 10.1109/TSMCC.2011.2161285.
- L. Marlina, M. lim, and A. P. Utama Siahaan, "Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms)," *Int. J. Eng. Trend Techno.*, 2016, doi: 10.14445/22315381/ijett-v38p268.
- "Decision Support System for Medical Diagnosis Using Data Mining," *Int. J. Comput. Sci. Issues*, 2011.
- B. HSSINA, A. MERBOUHA, H. EZZIKOURI, and ERRITALI, "A comparative study of decision tree ID3 and C4.5," *Int. J. Advan. Comp. Sci. Appl.*, 2014, doi: 10.14569/specialissue.2014.040203.
- C. J. Mantas and J. Abellán, "Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data," *Expert Syst. Appl.*, 2014, doi: 10.1016/j.eswa.2014.01.017.
- S. Angra and S. Ahuja, "Implementation of data mining algorithms on student's data using rapid miner," 2017, doi: 10.1109/ICBDACI.2017.8070869.
- K. M. Ting, "Confusion Matrix," in *Encyclopedia of Machine Learning and Data Mining*, 2017.
- A. Ridwan, P. N. Andono, and C. Supriyanto, "Optimasi Klasifikasi Status Gizi Balita Berdasarkan Indeks Antropometri Menggunakan Algoritma Naive," *Teknol. Inf.*, 2018.