

PENERAPAN ALGORITMA C4.5 UNTUK MEMBUAT MODEL PREDIKSI PASIEN YANG MENGIDAP PENYAKIT DIABETES

¹⁾ Sunanto, ²⁾ Ghazi Falah

^{1,2)}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau

^{1,2)}Jl. Tuanku Tambusai Ujung, Simpang Komersil Arengka, Kota Pekanbaru, Riau 28290

E-mail: sunanto@umri.ac.id

ABSTRAK

Diabetes adalah penyakit dimana pankreas tidak bisa memproduksi insulin dengan baik. Insulin merupakan hormon yang di hasilkan dari pankreas, yang berguna sebagai pintu untuk menyalurkan glukosa dari makan yang diserap untuk di alirkan ke dalam sel-sel darah agar tubuh dapat menghasilkan energi. Sedangkan menurut WHO diabetes adalah penyakit yang sangat mematikan yang menduduki peringkat 9 dunia. Penderita penyakit diabetes pada umumnya meninggal dunia dengan kondisi rusak nya beberapa organ vital seperti, jantung, ginjal dan hati. Para penderita awal penyakit diabetes tidak diketahui secara pasti sehingga penderita penyakit diabetes yang telah dirawat dalam kondisi yang memperhatikan. Untuk menekan jumlah kematian akibat penyakit diabetes dibutuhkan suatu sistem yang dapat mengidentifikasi gejala dini penyakit diabetes, agar penderita penyakit diabetes dapat ditangani langsung dengan baik. Teknologi data mining dapat membatu membangun suatu sistem untuk memprediksi penyakit diabetes menggunakan algoritma *Decision Tree C4.5*. pada penelitian ini data prediksi penyakit diabetes diambil dari UCI Repository. Kemudian data diolah dengan tahapan, select data, pre-processing dan split validation menggunakan RapidMiner. Hasil pengolahan data menggunakan RapidMiner, berupa rules yang dapat digunakan untuk memprediksi penyakit diabetes. Rules yang dihasilkan dari pohon keputusan RapidMiner memiliki akurasi 95, 51%

Kata Kunci: penderita, diabetes, insulin, data mining, RapidMiner

ABSTRACT

Diabetes is a disease in which the pancreas cannot produce insulin properly. Insulin is a hormone produced from the pancreas, which is useful as a door to channel glucose from food that is absorbed to be flowed into blood cells so that the body can produce energy. Meanwhile, according to WHO diabetes is a very deadly disease which is ranked 9th in the world. People with diabetes generally die from damage to several vital organs such as the heart, kidneys and liver. The early sufferers of diabetes are not known with certainty so that diabetics who have been treated are in a worrying condition. To reduce the number of deaths due to diabetes, a system is needed that can identify early symptoms of diabetes, so that people with diabetes can be handled properly. Data mining technology can help build a system to predict diabetes using the C4.5 Decision Tree algorithm. In this study, diabetes prediction data was taken from the UCI Repository. Then the data is processed in stages, select data, pre-processing and split validation using RapidMiner. The results of data processing using a RapidMiner, in the form of rules that can be used to predict diabetes. The rules generated from the RapidMiner Decision Tree have an accuracy of 95.51%.

Keyword: sufferers, diabetes, insulin, data mining, RapidMiner

PENDAHULUAN

Menurut WHO diabetes adalah penyakit paling mematikan peringkat 9 di dunia. Sedangkan menurut Kemenkes Indonesia merupakan peringkat 7 paling banyak menderita diabetes di dunia [1]. Diabetes Merupakan penyakit yang tidak menunjukkan gejala yang jelas sehingga membuat penderitanya tidak sadar terkena penyakit diabetes sehingga penyakit diabetes

biasanya baru di ketahui apabila telah merusak bagian vital seperti ginjal, mata dan saraf manusia. Penyakit diabetes juga menyebabkan orang yang menderitanya mengidap penyakit jantung, juga dapat menyebabkan kematian pada ibu yang akan melahirkan. Selain itu terdapat factor penularan penyakit diabetes kepada bayi yang akan lahir menurut [2] Diabetes adalah penyakit dimana pankreas tidak bisa memproduksi insulin dengan baik,

saat insulin yang digunakan oleh tubuh atau saat tubuh tidak bisa menghasilkan insulin atau tidak dapat menyalurkan insulin dengan baik oleh pankreas. Insulin merupakan salah satu hormon yang di hasilkan dari pankreas, yang bertugas sebagai pintu untuk menyalurkan glukosa dari makan yang diserap untuk di alirkan ke dalam sel-sel darah agar tubuh dapat menghasilkan energy untuk di gunakan [3]

Salah satu yang menyebabkan meningkatnya jumlah penderita diabetes adalah karena keterlambatan diagnosis penyakit diabetes itu sendiri. Sering terjadi pasien yang meninggal sebelum bisa di diagnosa menderita diabetes akibat komplikasi penyakit yang di sebabkan diabetes. Terdapat bermacam macam *variable* dan factor-faktor yang tidak disadari dalam diagnosa diabetes yang menyebabkan keterlambatan diagnosa. Diabetes akan berbahaya apabila kardar gula dalam darah sudah tinggi atau melebihi batas normal. Jika diabetes tidak di tangani dengan baik maka akan menimbulkan menimbulkan berbagai masalah lain seperti menyebabkan kompilkasi berbagai penyakit yang akan membahayakan nyawa pasien.

Menurut *International Diabetes Federation* (IDF) Jumlah penderita penyakit diabetes disetiap negara meningkat dan usia terbanyak orang dengan penyakit diabetes berada di usia antara 49-59 tahun. Pada kasus umum penderita penyakit diabetes akan menyebabkan resiko meningkatnya menderita penyakit cardio vascular atau sakit jantung sebesar dua sampai empat kali lebih banyak di bandingkan dengan yang tidak menderita penyakit diabetes. Selain dari cardio vascular dampak yang di akibatkan oleh penyakit diabetes seperti penyakit retinopati pada mata dan neuropati , oleh karena itu di butuhkan prediksi lebih awal untuk penyakit diabetes berdasarkan dari atribut

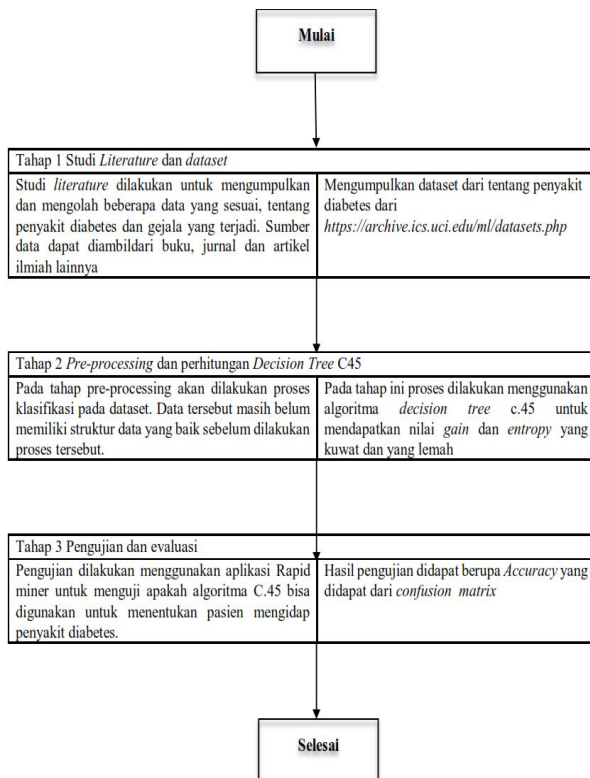
pendukung dari penyakit diabetes menurut [4] Salah satu cara untuk klasifikasi penyakit diabetes dengan menggunakan data mining.

Data mining di gunakan untuk melakukan proses ekstraksi informasi yang tersembunyi dari dataset yang banyak dan terdapat beberapa teknik dalam data mining seperti klasifikasi ,clustering, regresi dan asosiasi yang akan di gunakan dalam data pada bidang medis [5]. Pada data mining ini akan melakukan klasifikasi dimana akan memasukan berbagai data dan memasukannya pada kelas tertentu[6]. Di dalam melakukan prediksi data ini metode yang akan di gunakan adalah *Decision Tree* C45. Metode *Decision Tree* dengan algoritma C45 bisa melakukan prediksi dari berbagai informasi berdasarkan data yang digunakan menghitung kemungkinan terjadinya penyakit berdasarkan atributnya dapat digunakan dan juga untuk melihat seberapa efektifnya algoritma *Decision Treedi* gunakan untuk deteksi penyakit diabetes.

METODE

Tahapan penelitian yang dilakukan Metode yang di gunakan pada penelitian skripsi dilakukan seperti diagram pada gambar 1 memiliki 3 tahapan utama penelitian yaitu

- 1) Studi literature dan pengambilan dataset termasuk melakukan validasi data yang didapat menggunakan data pembanding yang lain;
- 2) Preprocessing dan perhitungan *Gain* dan nilai *entropy* menggunakan algoritma C45;
- 3) Pengujian dan hasil penelitian menggunakan aplikasi RapidMiner aplikasi ini dapat membuktikan apakah algoritma C45 dapat berjalan dengan baik untuk memprediksi pasien yang mengindap penyakit diabetes.

Gambar 1 *Framework* Penelitian

1. Pengumpulan data

Pada tahap pertama akan dilakukan pencarian referensi dari studi literature .Studi Literature adalah Mencari data mempelajari banyak data dari berbagai sumber buku, jurnal, modul, artikel baik perpustakaan maupun media internet yang berhubungan dengan penyakit diabetes [7]. Selajutnya berdasarkan referensi itu dapat mempelajari masalah yang ada untuk rancangan sistem , pencari studi literature berkaitan dengan topik data mining, decision tree C45 dan penyakit

2. Pengambilan data

Pengambilan data, data yang dikumpulkan yakni berapa banyak gejala yang di alami oleh penderita diabetes untuk mengklasifikasikannya ke dalam kesimpulan penyakit yang di diagnosis. Serta mengumpulkan data set tentang penyakit diabetes dari <https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset>.Pengola

DOI : <https://doi.org/10.36341/rabit.v7i2.2435>

han data dilakukan dengan menggunakan data mining dan menggunakan algoritma *Decision Tree C45*.

3. Pre-processing

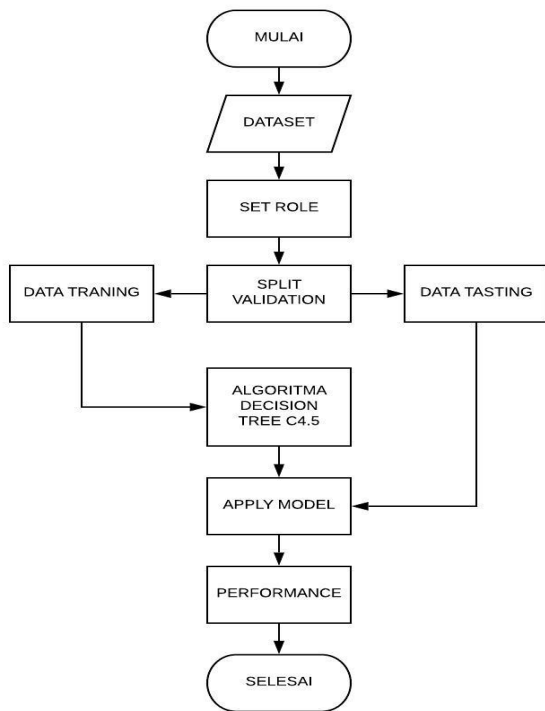
Tahap *preprocessing* data yang merupakan proses persiapan data sebelum data tersebut digunakan pada proses prediksi. Untuk membuat hasil prediksi yang baik, maka harus menggunakan data yang baik pula (lengkap, benar, konsisten, terintegrasi) [8]. Teknik preprocessing: Data preprocessing merupakan salah satu paling tugas data mining yang meliputi persiapan dan transformasi data ke dalam bentuk yang sesuai untuk pertambahan prosedur. Data preprocessing bertujuan untuk mengurangi data ukuran, menemukan hubungan antara data, data yang menormalkan, outliers menghapus dan fitur ekstrak data[9]. Selain itu preprocessing juga di gunakan untuk mengeliminasi varian-varian yang tidak di gunakan pada data[10].

4. Perhitungan *Decision Tree C4.5*

Tahap Menghitung *Decision Tree C4.5* dalam proses ini akan mencari data gain terbesar yang ada pada data dan akan di buat menjadi pohon keputusan sebagai acuan pengambilan data yang di proses. Dari perhitungan ini dapat pohon keputusan dari pohon keputusan ini kita mendapatkan rule yang akan kita gunakan dalam memprediksi data yang ada.

5. Pengujian

Pada sistem yang akan dilakuakan untuk menentukan nilai confudion matrix sebagai hasil penelitian adalah dengan alur Set Role menjelaskan bagaimana operator menjalankan atribut yang di tentukan. Atribut yang di akan di tentukan set role adalah atribut label yang akan di gunakan sebagai hasil dari klasifikasi sistem yang di jalankan [11]. Alur sistem pengujian yang ditampilkan pada gambar 2.



Gambar 2 Alur Sistem

Tahapan pengujian selanjutnya adalah Split validation merupakan teknik yang digunakan untuk membagi dataset menjadi data training dan data testing berdasarkan rasio yang ditentukan. Split Validation akan membagi dataset pada rasio yang dimasukkan oleh pengguna aplikasi. Split Validation hanya melakukan validasi satu kali [12]. Sebaiknya rasio yang pada data testing lebih besar dari pada data tasting pada data ini perbandingan yang digunakan adalah 70:30, 70 untuk data testing dan 30 untuk data training [13]. Setelah data di bagi dua , setelah itu data yang dibagi dua data testing akan di klasifikasikan menggunakan *Decision Tree C4.5* untuk mengetahui tingkat akurasi data yang di uji. Apply Model membuat prediksi menggunakan data testing yang sudah dalam model *Decision Tree C4.5* dan di uji dengan data training. Performance operator menghitung tingkat kesalahan dan akurasi prediksi dan akhirnya operator akan menghasilkan akurasi berdasarkan confusion matrix [14]. Confusion

matrix merupakan metode yang di gunakan untuk menghitung akurasi pada data mining. Confusion Matrix akan menghasilkan recall, precision, accuracy dan error rate [8].Rumus untuk menghitung tingkat akurasi pada *Confusion Matrix* dengan deskripsi variabel sebagai berikut :

- a. TP = True Positif
- b. TN = True Negatif
- c. FP= False Positif
- d. FN = False Negatif

1. Recall= $TP / (FN+TP)$
2. Precision= $TP / (FP+TP)$
3. Accuracy= $(TP+TN) / (TP+FN+FP+TN)$
4. Error rate= $(FN+FP) / (TP+FN+FP+TN)$

Jika dalam pengujian sudah mendapatkan hasil, maka tahapan selanjutnya adalah mempersentasikan hasil dari penelitian yang di sertai dengan dokumen laporan penelitian.

HASIL

Pembahasan ini meliputi analisis data, preprocessing, perhitungan pohon keputusan, pengujian dataset dengan algoritma *Decision Tree C4.5* dan validasi menggunakan RapidMiner, yang mana nantinya akan membantu dalam menyelesaikan laporan penelitian ini.

1. Analisis Data

Pada analisis data ini melihat hubungan antara variable yang satu dan variable yang lain[15].Pada penelitian ini data yang digunakan adalah data public UCI Repository Atribut data set yang di gunakan pada penelitian ini Usia, Jenis Kelamin, Sering buang air kecil, Sering haus, penurunan berat badan secara tiba-tiba, lemas ,Banyak makan, Genital thrush, Pandangan kabur, gatal-gatal,

mudah marah, Luka susah sembuh, partial paresis, muscle stiffness, kebotakan, kegemukan, dan class, pada tabel usia menampilkan numeric, pada tabel class menampilkan positive dan negative, jumlah data training yang digunakan sebanyak 520 data.

2. Pre-processing

Pada tahap preprocessing akan melakukan klasifikasi pada dataset. Data tersebut masih belum memiliki struktur yang baik atau diperlukan, dikarenakan terdapat numeric. Agar data dapat digunakan, maka harus dilakukan beberapa tahapan preprocessing, namun selain itu terdapat tahapan lain yang dapat dilakukan diluar preprocessing sesuai kebutuhan.

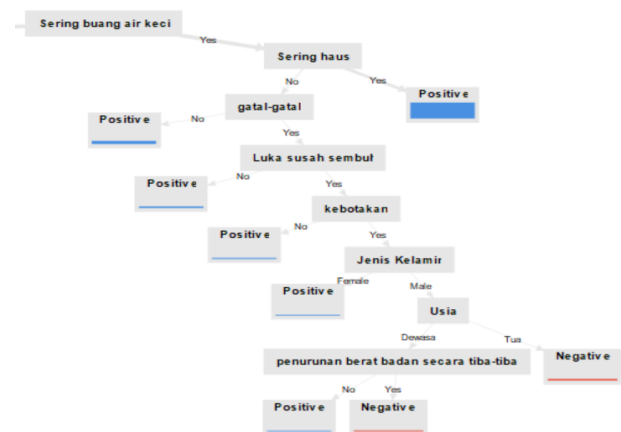
3. Perhitungan

Pada perhitungan ini menggunakan algoritma *Decision Tree* C4.5 untuk menentukan cabang dalam *Decision Tree* menggunakan gain terbesar yang ada sebagai cabang utama dalam pohon keputusan adapun tahapan perhitungan yang dilakukan dimulai dari cabang pertama perhitungan tersebut.

- 1) Perhitungan cabang Ke-1 menghasilkan nilai gain tertinggi 0,362251 pada “**sering buang air kecil**” pada cabang ini memiliki potensi 2 cabang yang akan ditelusuri kembali yaitu positif dan negatif.
- 2) Perhitungan Pada cabang Ke-2 ini di dapatkan nilai gain tertinggi 0,8638179 pada “**Sering haus**” yang akan ditelusuri pada dua cabang positif dan negatif;
- 3) Perhitungan Pada cabang Ke-3 ini di dapatkan nilai gain tertinggi 0,894923172 pada “**gatal gatal**” yang akan ditelusuri pada dua cabang positif dan negatif;
- 4) Perhitungan pada cabang Ke-4 Pada cabang ini di dapatkan nilai gain tertinggi 0,9171858 pada “**luka susah sembuh**” yang akan ditelusuri pada dua cabang

positif dan negatif;

- 5) Perhitungan pada cabang Ke-5 Pada cabang ini di dapatkan nilai gain tertinggi 0,934107 pada “**luka susah sembuh**” yang akan ditelusuri pada dua cabang positif dan negatif;
- 6) Perhitungan pada cabang Ke-6 Pada cabang ini di dapatkan nilai gain tertinggi 0,9441529 pada “**jenis kelamin**” potensi cabang nya adalah pria dan wanita;
- 7) Perhitungan pada cabang Ke-7 Pada cabang ini di dapatkan nilai gain tertinggi 0,9506409 pada “**usia**” yang akan ditelusuri pada dua cabang positif dan negatif;
- 8) Perhitungan pada cabang Ke-8 Pada cabang ini di dapatkan nilai gain tertinggi 0,9612366 pada “**penurunan berat badan secara tiba tiba**” yang akan ditelusuri pada dua cabang positif dan negatif.
- 9) Perhitungan dari cabang Ke-1 hingga cabang Ke-8 maka didapatkan pohon keputusan yang disajikan pada gambar 3



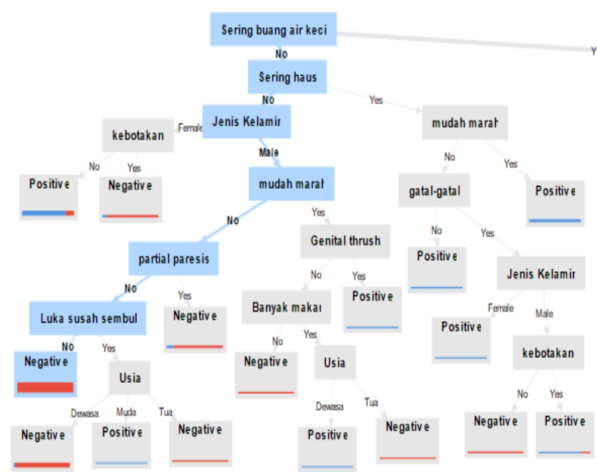
Gambar 3 *Decision Tree* cabang pertama

- 10) Perhitungan pada cabang Ke-9 Pada cabang ini di dapatkan nilai gain tertinggi 0,5952022 pada “**sering haus**” yang akan ditelusuri pada dua cabang positif dan negatif;
- 11) Perhitungan pada cabang Ke-10 Pada cabang ini di dapatkan nilai gain tertinggi

- 0,921228 pada “**mudah marah**” yang akan ditelusuri pada dua cabang positif dan negatif;
- 12) Perhitungan pada cabang Ke-11 Pada cabang ini di dapatkan nilai gain tertinggi 0,930467 pada “**gatal gatal**” yang akan ditelusuri pada dua cabang positif dan negatif;
- 13) Perhitungan pada cabang Ke-12 Pada cabang ini di dapatkan nilai gain tertinggi 0,940045 pada “**jenis kelamin**” pria atau wanita.
- 14) Perhitungan pada cabang Ke-13 Pada cabang ini di dapatkan nilai gain tertinggi 0,954295 pada “**kebotakan**” yang akan ditelusuri pada dua cabang positif dan negatif;
- 15) Perhitungan pada cabang Ke-14 Pada cabang ini di dapatkan nilai gain tertinggi 0,740212 pada “**jenis kelamin**” Pria atau wanita
- 16) Perhitungan pada cabang Ke-15 Pada cabang ini di dapatkan nilai gain tertinggi 0,905742 pada “**kebotakan**” yang akan ditelusuri pada dua cabang positif dan negatif ;
- 17) Perhitungan pada cabang Ke-16 Pada cabang ini di dapatkan nilai gain tertinggi 0,857636 pada “**mudah marah**” yang akan ditelusuri pada dua cabang positif dan negatif ;
- 18) Perhitungan pada cabang Ke-17 Pada cabang ini di dapatkan nilai gain tertinggi 0,945752 pada “**genital thrush**” yang akan ditelusuri pada dua cabang positif dan negative;
- 19) Perhitungan pada cabang Ke-18 Pada cabang ini di dapatkan nilai gain tertinggi 0,950641 pada “**banyak makan**” yang akan ditelusuri pada dua cabang positif dan negatif ;
- 20) Perhitungan pada cabang Ke-19 Pada cabang ini di dapatkan nilai gain tertinggi

0,961237 pada “**usia**” tua dan muda .

- 21) Perhitungan pada cabang Ke-20 Pada cabang ini di dapatkan nilai gain tertinggi 0,894362 pada “**partial perisis**” yang akan ditelusuri pada dua cabang positif dan negatif ;
- 22) Perhitungan pada cabang Ke-21 Pada cabang ini di dapatkan nilai gain tertinggi 0,923532 pada “**Luka susah sembuh**” yang akan ditelusuri pada dua cabang positif dan negatif;
- 23) Perhitungan pada cabang Ke-22 Pada cabang ini di dapatkan nilai gain tertinggi 0,948482 pada “**usia**” yang dimana memiliki 3 cabang yang akan terjadi.
- 24) Dari cabang ke 11 sampai cabang ke 22 menghasilkan *Decision Tree* pada cabang ke kedua



Gambar 4 *Decision Tree* cabang kedua

Rule yang sudah di hitung di atas akan terlihat seperti berikut:

- Sering buang air kecil = No
- | Sering haus = No
- | | Jenis Kelamin = Female
- | | | kebotakan = No: Positive {Positive=31, Negative=6} -Rule 1
- | | | kebotakan = Yes: Negative {Positive=1, Negative=13} -Rule 2
- | | Jenis Kelamin = Male
- | | | mudah marah = No
- | | | | partial paresis = No
- | | | | Luka susah sembuh = No: Negative {Positive=0, Negative=91} -Rule 3

Tahapan Penggunaan proses pada gambar 3 RapidMiner Model.

- Memasukan/*Retrieve* data diabetes yang dijadikan sebagai model untuk memprediksi penyakit diabetes ditetapkan sebagai data sample
- Membuat Set Role adalah menetapkan atribut dan peran yang digunakan pada data diabetes.
- Melakukan Validation dalam hal ini split validation memisahkan data uji dan data latih
- Memilih algoritma *Decision Tree* untuk menetapkan pengolahan data
- Lakukan Apply Model/Penerapan model *Decision Tree*
- Performance melihat/menampilkan performa dari hasil presentase pada proses pengolahan data menggunakan Algoritma *Decision Tree C.45*.

accuracy: 95.51%

	true Positive	true Negative	class precision
pred. Positive	92	3	96.84%
pred. Negative	4	57	93.44%
class recall	95.83%	95.00%	

Gambar 7 Hasil Pengujian

Dari hasil perhitungan Confusion Matrix yang sudah dilakukan mendapatkan hasil class recall true positive sebesar 95,83% dan true negative sebesar 95% . hasil dari class precision sebesar pred positive sebesar 96,84% dan pred negative sebesar 93,44 dan error clasifikation sebesar 4,49% dan akurasi prediksi sebesar 95,51%.

KESIMPULAN

Kesimpulan yang di dapatkan dari hasil penelitian pengujian yang dilakakukan terhadap sistem prediksi penyakit diabetes menggunakan metode algoritma *Decision Treec4.5* adalah penggunaan algoritma

Decision TreeC4.5 untuk mengklasifikasikan penyakit diabetes merupakan salah satu cara untuk memprediksi penyakit diabetes yang bagus untuk di gunakan untuk klasifikasi awal penyakit diabetes. Prediksi penyakit menggunakan algoritma *Decision TreeC4.5* memiliki hasil yang bagus dapat dilihat dari hasil perhitungan confusion matrix yang mendapatkan hasil accuracy sebesar 95,51% dan juga hanya mendapatkan error classification sebesar 4,49%.

DAFTAR PUSTAKA

- [1] S. Pangribowo, "Infodatin-2020-Diabetes-Melitus.pdf." 2020.
- [2] M. B. Hanif and Khoirudin, "Sistem Aplikasi Prediksi Penyakit diabetes Menggunakan Fiture Selection Korelasi Pearson dan Klasifikasi Naïve Bayes," *Pengemb. Rekayasa dan Teknol.*, vol. 16, no. 2, pp. 199–205, 2020.
- [3] A. Ridwan, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020, doi: 10.47970/siskom-kb.v4i1.169.
- [4] Putri, E. Ucha, Sanni Irawan, and F. Rizky, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes," *KESATRIA(J. penerapan Sist. Inf. dan Manaj.*, vol. 2, no. 1, pp. 39–46, 2021.
- [5] Y. A. Fiandra, S. Defit, and Y. Yuhandri, "Penerapan Algoritma C4.5 untuk Klasifikasi Data Rekam Medis berdasarkan International Classification Diseases (ICD-10)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 2, pp. 82–89, 2017, doi: 10.29207/resti.v1i2.48.
- [6] I. Lishania, R. Goejantoro, and Y. N. Nasution, "Perbandingan Klasifikasi Metode Naive Bayes dan Metode

- Decision Tree* Algoritma (J48) pada Pasien Penderita Penyakit Stroke di RSUD Abdul Wahab Sjahranie Samarinda,” *J. Eksponensial*, vol. 10, no. 2, pp. 135–142, 2019.
- [7] A. Mujahidin and D. Pribadi, “Penerapan algoritma C4 . 5 untuk diagnosa penyakit pneumonia pada anak balita berbasis mobile,” *J. Swabumi*, vol. 5, no. 2, pp. 155–161, 2017, [Online]. Available: <https://ejournal.bsi.ac.id/ejurnal/index.php/swabumi/article/view/2523>.
- [8] E. P. K. Orpa, E. F. Ripanti, and T. Tursina, “Model Prediksi Awal Masa Studi Mahasiswa Menggunakan Algoritma *Decision Tree* C4.5,” *J. Sist. dan Teknol. Inf.*, vol. 7, no. 4, p. 272, 2019, doi: 10.26418/justin.v7i4.33163.
- [9] S. A. Alasadi and W. S. Bhaya, “Review of data preprocessing techniques in data mining,” *J. Eng. Appl. Sci.*, vol. 12, no. 16, pp. 4102–4107, 2017, doi: 10.3923/jeasci.2017.4102.4107.
- [10] J. P. Fortin, T. J. Triche, and K. D. Hansen, “Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi,” *Bioinformatics*, vol. 33, no. 4, pp. 558–560, 2017, doi: 10.1093/bioinformatics/btw691.
- [11] A. Massaro, V. Maritati, and A. Galiano, “Data Mining Model Performance of Sales Predictive Algorithms Based on Rapidminer Workflows,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 10, no. 3, pp. 39–56, 2018, doi: 10.5121/ijcsit.2018.10303.
- [12] G. Mahajan, B. Saini, and T. Almas, “Taxonomy on RapidMiner Using Machine Learning,” *SSRN Electron. J.*, pp. 2410–2415, 2019, doi: 10.2139/ssrn.3363071.
- [13] S. R. Phandinata, E. T. Atmodiwirjo, and D. Basaria, “Developmental Individual-Differences Relationship-Based (Dir) Floortime Dalam Meningkatkan Komunikasi Dua Arah Pada Kasus Autism Spectrum Disorder (Asd),” *Psibernetika*, vol. 10, no. 2, pp. 103–113, 2017, doi: 10.30813/psibernetika.v10i2.1046.
- [14] U. Celik and C. Basarir, “The Prediction of Precious Metal Prices via Artificial Neural Network by Using RapidMiner,” *Alphanumeric J.*, vol. 5, no. 1, pp. 45–45, 2017, doi: 10.17093/alphanumeric.290381.
- [15] A. Purwanto, M. Asbari, and T. I. Santoso, “Analisis Data Penelitian Marketing: Perbandingan Hasil antara Amos, SmartPLS, WarpPLS, dan SPSS untuk Jumlah Sampel Besar,” *J. Ind. Eng. Manag. Res.*, vol. 2, no. 4, pp. 216–227, 2021, [Online]. Available: <https://www.jiemar.org/index.php/jiemar/article/view/178/138>.