

## Analisa Classification Decision Tree C45 dan Naïve Bayes Pada Indikasi Penyakit Diabetes Menggunakan Rapid Miner

Iswadi Hamzah<sup>1</sup>, Zulham Sitorus<sup>2</sup>, Khairul<sup>3</sup>

<sup>1</sup> Magister Teknologi Informasi,  
Universitas Pembangunan Panca Budi, Kota Medan, Indonesia  
E-mail: [iswadi.unpab@gmail.com](mailto:iswadi.unpab@gmail.com),

Corresponding Author: Iswadi Hamzah

---

### ABSTRACT

*In Indonesia, the rate of diabetes sufferers continues to increase, so this is deemed necessary to pay attention to by the Indonesian people in particular, for this reason this research is not the first to be conducted. Predicting diabetes can be done using various methods through various algorithms which are quite diverse, therefore it is necessary to conduct research on the algorithms used. To obtain new information, the Decision Tree algorithm with Naïve Bayes was tested using the Rapid Miner application. This test is carried out on data that has the attribute HighBP, HighChol, CholCheck, BMI, Smoker, Stroke, Heart Diseaseor Attack, Phys Activity, Fruits, Veggies, HvyAlcoholConsump, AnyHealthcare, NoDocbcCos, GenHlth, MentHlth, PhysHlth, DiffWalk, Sex, Age, Education, Income. All of these attributes serve as a guide in determining results, so that it can be known that the patient has diabetes.*

**Keywords:** *Decision Tree, Diabetes, Naïve Bayes, Rapid Miner*

### ABSTRAK

Di Indonesia, tingkat penderita diabetes terus meningkat, sehingga hal ini dipandang perlu untuk dijadikan perhatian oleh masyarakat Indonesia khususnya, penelitian ini bukanlah merupakan penelitian yang pertama kali dilakukan. Didalam melakukan prediksi penyakit diabetes dapat dilakukan dengan berbagai metode melalui berbagai algoritma yang cukup beragam, oleh sebab itu perlu dilakukannya penelitian terhadap algoritma yang digunakan. Untuk mendapatkan informasi yang baru, maka dilakukan pengujian terhadap algoritma Decision Tree dengan Naïve Bayes dengan menggunakan aplikasi Rapid Miner. Pengujian ini dilakukan pada data yang memiliki atribut HighBP, HighChol, CholCheck, BMI, Smoker, Stroke, Heart Diseaseor Attack, Phys Activity, Fruits, Veggies, HvyAlcoholConsump, AnyHealthcare, NoDocbcCos, GenHlth, MentHlth, PhysHlth, DiffWalk, Sex, Age, Education, Income. Keseluruhan atribut ini menjadi pedoman dalam menentukan hasil, sehingga dapat diketahui bahwa penderita mengalami penyakit diabetes.

**Kata Kunci:** *Decision Tree, Diabetes, Naïve Bayes, Rapid Miner*

---

### 1. Pendahuluan

Data Mining adalah proses penggalian informasi dan pola yang bermanfaat dari suatu data yang sangat besar. Proses data mining terdiri dari pengumpulan data, ekstraksi data, analisa data, dan statistik data. Ia juga umum dikenal sebagai knowledge discovery, knowledge extraction, data/pattern analysis, information harvesting, dan lainnya. [1] Decision tree atau pohon keputusan merupakan metode klasifikasi yang sangat kuat dan populer. Decision tree dapat mengubah data yang sangat besar menjadi sebuah pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami, karenanya decision tree merupakan teknik klasifikasi yang mudah untuk dipelajari dan sangat populer digunakan. Decision tree merupakan teknik yang digunakan untuk melakukan klasifikasi terhadap sekumpulan objek atau data dengan representasi struktur pohon. [2] Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes yaitu dengan memprediksi masa depan berdasarkan data dari masa lalu.



Naïve Bayes untuk setiap kelas keputusan menghitung probabilitas dengan syarat bahwa kelas keputusan adaah benar, mengingat vector informasi objek. Algoritma ini mengasumsikan bahwa atribut objek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari “master” tabel keputusan. [3]

Berdasarkan data dan informasi yang dianalisa, rumusan masalah pada penelitian ini adalah, bagaimana memodelkan dan menguji dataset diabaetes, menggunakan model decision tree serta naïve bayes untuk melakukan perbandingan nilai akurasi dari kedua model tersebut dalam mengidentifikasi penyebab penyakit diabetes. Penelitian ini menggunakan data set dari kaggle dengan asumsi data training 80% dan data testing 20% yang akan dilakukan pada penggunaan pelatihan dan pengujian dengan menerapkan kedua metode yaitu algoritma decision tree dan naïve bayes dalam melakukan penentuan faktor penyebab penyakit diabetes. Rapid miner merupakan platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan yang bernama sama dengan pihak yang mmenyediakan lingkungan terintegrasi untuk persiapan data, penambangan teks, dan analisis prediktif. Rapid miner dilengkapi dengan satu prosesor logika dan 10.000 baris data. Rapid miner yang tersedia berada di bawah lisensi AGPL. Data penelitian ini bersumber dari dataset yang tersedia pada Kaggle memiliki 22 atribut dan 236.378 record yang telah di bagi menjadi 2 buah data, yaitu data training sejumlah 80% dan untuk data testing digunakan sejumlah 20%.

Table 1 : Dataset Diabetes Indikator

Diabetes_binary	HighBP	HighChol	Chol/Check	BMI	Smoker	Stroke	HeartDiseaseorAttack	PhysActivity	Fruits	Veggies	HvyAlcoholConsump	AnyHealthcare	NoDocbcCost	GenHlth	MentHlth	PhysHlth	DiffWalk	Sex	Age	Education	Income
00.00	0	01.00	1	15.00	01.00	00.00	00.00	0	1	1	0	1	00.00	05.00	10.00	20.00	00.00	0	11	04.00	05.00
01.00	1	00.00	1	28.00.00	00.00	00.00	01.00	0	1	0	0	1	00.00	02.00	00.00	00.00	00.00	0	11	04.00	03.00
01.00	1	01.00	1	33.00.00	00.00	00.00	00.00	1	1	1	0	1	00.00	02.00	10.00	00.00	00.00	0	9	04.00	07.00
01.00	0	01.00	1	29.00.00	00.00	01.00	01.00	1	1	1	0	1	00.00	05.00	00.00	30.00.00	01.00	1	12	03.00	04.00
00.00	0	00.00	1	24.00.00	01.00	00.00	00.00	0	0	0	0	1	00.00	03.00	00.00	00.00	01.00	1	13	05.00	06.00
00.00	1	00.00	1	40.00.00	01.00	00.00	00.00	1	1	1	0	1	00.00	03.00	05.00	25.00.00	01.00	0	10	04.00	08.00
00.00	1	01.00	1	27.00.00	01.00	00.00	00.00	0	0	1	0	1	00.00	04.00	25.00.00	00.00	00.00	0	10	05.00	03.00
01.00	0	00.00	1	24.00.00	00.00	00.00	01.00	0	0	1	0	1	00.00	04.00	00.00	00.00	00.00	1	12	06.00	07.00
00.00	1	01.00	1	30.00.00	00.00	00.00	00.00	0	1	1	0	1	00.00	02.00	00.00	00.00	00.00	0	7	04.00	06.00
00.00	1	01.00	1	36.00.00	01.00	00.00	00.00	0	0	0	0	1	00.00	04.00	00.00	00.00	00.00	1	10	04.00	08.00
01.00	0	00.00	1	33.00.00	01.00	00.00	00.00	1	0	1	0	1	00.00	04.00	00.00	00.00	00.00	1	6	05.00	02.00
01.00	1	01.00	1	30.00.00	00.00	00.00	00.00	0	1	1	0	1	00.00	03.00	02.00	04.00	00.00	0	11	05.00	04.00
01.00	0	00.00	1	27.00.00	00.00	00.00	00.00	0	0	1	0	1	00.00	04.00	00.00	30.00.00	00.00	1	13	06.00	05.00
01.00	0	00.00	1	31.00.00	01.00	00.00	00.00	0	1	1	0	1	00.00	03.00	10.00	10.00	00.00	0	8	04.00	08.00
00.00	1	01.00	1	27.00.00	01.00	00.00	00.00	0	1	1	0	1	00.00	03.00	00.00	02.00	01.00	0	13	06.00	06.00
00.00	1	01.00	1	35.00.00	01.00	00.00	00.00	1	1	0	0	1	00.00	03.00	05.00	03.00	00.00	0	11	04.00	05.00
00.00	1	01.00	1	30.00.00	00.00	00.00	00.00	1	1	0	0	1	00.00	02.00	00.00	10.00	00.00	0	10	06.00	08.00
00.00	0	00.00	1	31.00.00	00.00	00.00	00.00	1	1	1	0	1	00.00	01.00	00.00	02.00	00.00	1	11	06.00	08.00
00.00	1	01.00	1	33.00.00	00.00	00.00	00.00	0	1	1	0	1	00.00	02.00	00.00	00.00	00.00	1	11	05.00	05.00
00.00	0	00.00	1	31.00.00	01.00	00.00	00.00	1	1	1	0	1	00.00	04.00	05.00	00.00	00.00	0	10	05.00	06.00
00.00	1	00.00	1	29.00.00	00.00	00.00	00.00	1	1	1	0	1	00.00	04.00	00.00	00.00	00.00	1	11	04.00	05.00
00.00	1	00.00	1	27.00.00	01.00	01.00	00.00	1	0	0	0	1	00.00	03.00	07.00	15.00	00.00	0	7	05.00	05.00
01.00	1	01.00	1	22.00	00.00	00.00	01.00	1	0	1	0	1	00.00	03.00	00.00	00.00	00.00	0	12	04.00	05.00
00.00	1	00.00	1	29.00.00	01.00	00.00	00.00	1	1	1	0	1	00.00	03.00	00.00	30.00.00	00.00	1	12	06.00	08.00
00.00	0	00.00	1	22.00	00.00	00.00	00.00	1	1	1	0	1	00.00	01.00	00.00	00.00	00.00	0	9	06.00	08.00
00.00	1	01.00	1	26.00.00	01.00	00.00	00.00	0	1	1	0	1	00.00	03.00	00.00	00.00	01.00	0	13	05.00	08.00
00.00	1	01.00	1	26.00.00	01.00	00.00	00.00	0	1	1	0	1	00.00	03.00	00.00	00.00	01.00	0	12	05.00	02.00
01.00	0	01.00	1	28.00.00	01.00	00.00	00.00	1	0	0	0	1	00.00	03.00	00.00	00.00	00.00	1	11	05.00	09.00
00.00	0	00.00	1	28.00.00	00.00	00.00	00.00	1	1	1	0	1	00.00	02.00	00.00	00.00	00.00	1	8	06.00	10.00
00.00	1	01.00	1	30.00.00	01.00	00.00	00.00	0	1	1	0	1	01.00	04.00	30.00.00	15.00	01.00	0	8	05.00	07.00



## Metodologi

Berisi mengenai tahapan-tahapan penelitian dimulai dari penelitian awal berupa pengumpulan data sampai kepada penerapan algoritma kepada aplikasi rapid miner sehingga ditemukan kecocokan antara hasil akhir terhadap masing-masing perhitungan. Output dari penelitian ini adalah analisa terhadap hasil prediksi faktor seseorang bisa terkena penyakit diabetes.

Pada penelitian ini, peneliti menemukan beberapa keadaan yang membuat peneliti tertarik untuk melakukan suatu penganalisaan terhadap kemungkinan – kemungkinan yang akan membuat seseorang terkena penyakit diabetes. Tahapan dari penelitian ini dilakukan dengan beberapa metode tahapan yang saling berhubungan antara satu dengan yang lainnya. Metode yang dilakukan berupa:

1. Mengambil dataset dari Kaggle dalam format csv.
2. Membagi dan memisahkan data training dan data testing dengan rasio 80:20 %.
3. Melakukan pengolahan data dengan menggunakan software rapid miner.
4. Setelah data diklasifikasikan ke dalam .csv, maka data tersebut kemudian akan diproses dengan menggunakan software rapid miner. Dengan menggunakan software tersebut maka data akan dapat dianalisa untuk mendapatkan sebuah keputusan.

KDD (Knowledge Discovery In Database) Istilah data mining dan knowledge discovery in database (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. [4] Proses KDD secara garis besar sebagai berikut :

1. Data Selection Pemilihan (Seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.
2. Pre Processing / Pembersihan Data Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi fokus KDD.
3. Transformation pada data yang telah dipilih, sehingga data tersebut sesuai proses data mining.
4. Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu
5. Evaluasi Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan

### a. Decision Tree C45

Decision tree adalah salah satu metode klasifikasi yang paling populer dan banyak digunakan karena bersifat praktis. salah satu metode pohon keputusan yang paling banyak digunakan adalah Iterative Dychotomizer Version 3. Metode ID3 berusaha membangun model klasifikasi yang berupa pohon keputusan secara top-down. Caranya adalah dengan mengevaluasi semua atribut menggunakan suatu ukuran statistik, biasanya berupa information gain, untuk mengukur efektifitas suatu atribut dalam mengklasifikasikan himpunan sampel data. Dengan tujuan untuk mendefinisikan gain, pertama-tama digunakanlah ide dari teori informasi yang disebut entropy. Entropy mengukur jumlah dari informasi yang ada pada atribut dengan rumus :

$$Entropy(S) = -P_+ \log_2 P_+ - P_- \log_2 P_-$$

S = ruang data sample

P<sub>+</sub> = jumlah bersolusi positif (mendukung)

P<sub>-</sub> = jumlah bersolusi negatif (tidak mendukung)



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.



Dari rumus entropy diatas dapat disimpulkan bahwa definisi entropy (S) adalah jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau-) dari sejumlah data acak pada suatu ruang sampel S. Entropy bisa dikatakan sebagai kebutuhan bit untuk meyakinkan suatu kelas. Semakin kecil nilai entropy maka semakin baik digunakan dalam mengekstraksi suatu kelas. Pada algoritma ID3 pengurangan entropy disebut dengan informasi gain. Pembagian sample S terhadap atribut A dapat dihitung information gain dengan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{v \in \text{nilai}(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Pilih atribut yang memiliki nilai information gain terbesar, ulangi proses perhitungan information gain akan terus dilaksanakan sampai semua data telah masuk dalam kelas yang sama. Atribut yang telah dipilih tidak diikuti lagi dalam perhitungan nilai information gain. [5]

### b. Naïve Bayes

Naive Bayes merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik. Naïve Bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. [6]. Teorema Bayes memiliki bentuk umum sebagai berikut :

$$P(H|X) = \frac{P(X|H)P(H)}{P(H)} \dots\dots\dots(1)$$

Di mana :

X : Kelas data yang belum diketahui

H : Hipotesa data X adalah kelas spesifik

P(H|X) : Kemungkinan Hipotesa H berdasarkan keadaan X (posteriori prob.)

P(H) : Kemungkinan Hipotesa H (prior prob.)

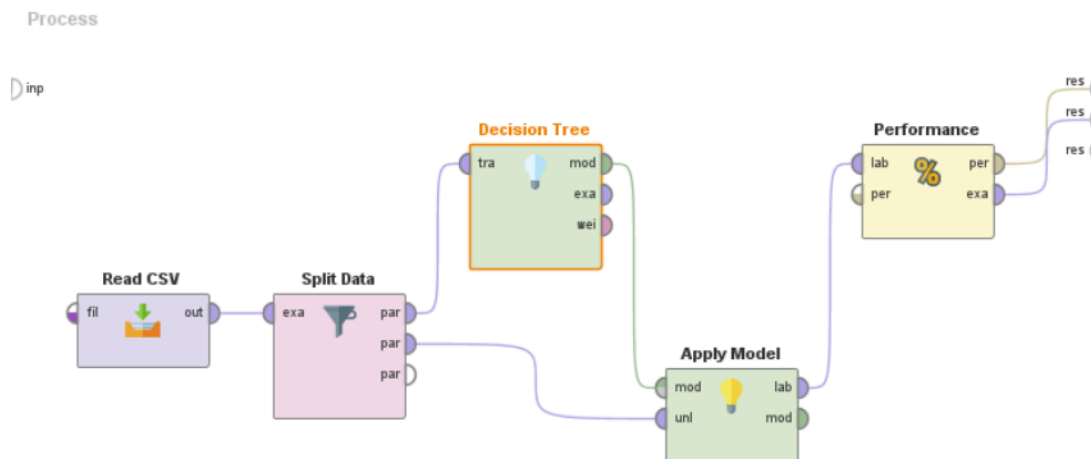
P(X|H) : Kemungkinan X berdasarkan keadaan tersebut

P(X) : Kemungkinan dari X

## Hasil dan Pembahasan

### a. Pengujian Decision Tree

Pengujian Analisa ini menggunakan Rapidminer seperti berikut ini :



Gambar 1 : Pengujian Decision Tree

Data Training yang digunakan dalam penelitian ini adalah data indikator diabetes sebanyak 47.276 dari 6.714 dinyatakan terindikasi diabetes dan 40.562 negatif diabetes.



Gambar 3 : Pohon Keputusan

Dengan hasil sebagai berikut :

```

BMI > 98.500: 1.0 {0.0=0, 1.0=2}
BMI ≤ 98.500
|   GenHlth > 3.500
|   |   Age > 3.500
|   |   |   BMI > 12.500
|   |   |   |   CholCheck > 0.500: 0.0 {0.0=17354, 1.0=9547}
|   |   |   |   CholCheck ≤ 0.500
|   |   |   |   |   Education > 1.500
|   |   |   |   |   |   Stroke > 0.500
|   |   |   |   |   |   |   HighChol > 0.500
|   |   |   |   |   |   |   |   Age > 11: 0.0 {0.0=3, 1.0=0}
|   |   |   |   |   |   |   |   Age ≤ 11: 1.0 {0.0=3, 1.0=7}
|   |   |   |   |   |   |   |   HighChol ≤ 0.500: 0.0 {0.0=17, 1.0=1}
|   |   |   |   |   |   |   |   Stroke ≤ 0.500
|   |   |   |   |   |   |   |   |   HeartDiseaseorAttack > 0.500
|   |   |   |   |   |   |   |   |   |   BMI > 40.500: 1.0 {0.0=1, 1.0=2}
|   |   |   |   |   |   |   |   |   |   BMI ≤ 40.500: 0.0 {0.0=39, 1.0=8}
|   |   |   |   |   |   |   |   |   |   HeartDiseaseorAttack ≤ 0.500: 0.0 {0.0=422, 1.0=29}
|   |   |   |   |   |   |   |   |   |   Education ≤ 1.500: 0.0 {0.0=1, 1.0=1}
|   |   |   |   |   |   |   |   |   |   BMI ≤ 12.500: 0.0 {0.0=2, 1.0=0}
|   |   |   |   |   |   |   |   |   |   Age ≤ 3.500
|   |   |   |   |   |   |   |   |   |   |   BMI > 82: 0.0 {0.0=1, 1.0=1}
|   |   |   |   |   |   |   |   |   |   |   BMI ≤ 82
|   |   |   |   |   |   |   |   |   |   |   |   BMI > 62.500
|   |   |   |   |   |   |   |   |   |   |   |   |   PhysHlth > 6
|   |   |   |   |   |   |   |   |   |   |   |   |   |   GenHlth > 4.500: 0.0 {0.0=2, 1.0=0}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   GenHlth ≤ 4.500
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   HighBP > 0.500: 1.0 {0.0=0, 1.0=3}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   HighBP ≤ 0.500: 0.0 {0.0=1, 1.0=1}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   PhysHlth ≤ 6: 0.0 {0.0=4, 1.0=0}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   BMI ≤ 62.500
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   HighChol > 0.500
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   BMI > 54.500
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   BMI > 58: 0.0 {0.0=4, 1.0=1}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   BMI ≤ 58: 1.0 {0.0=0, 1.0=3}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   BMI ≤ 54.500
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Stroke > 0.500
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   PhysHlth > 25.500: 1.0 {0.0=0, 1.0=2}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   PhysHlth ≤ 25.500: 0.0 {0.0=6, 1.0=3}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Stroke ≤ 0.500: 0.0 {0.0=335, 1.0=71}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   HighChol ≤ 0.500
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   BMI > 48.500
    
```



**Lisensi**  
 Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.



Gambar 4 : Pengujian Naïve Bayes

### SimpleDistribution

Distribution model for label attribute Diabetes\_binary

Class 0.0 (0.858)  
 21 distributions

Class 1.0 (0.142)  
 21 distributions

### Evaluasi dan Pengujian Hasil

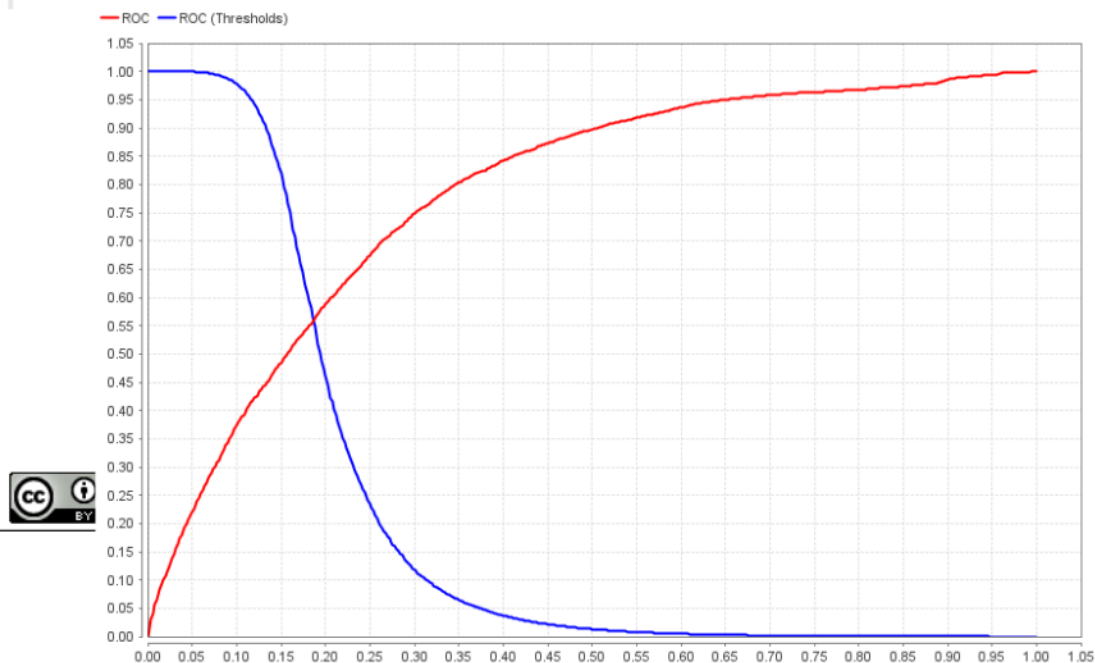
#### a. Decision Tree

Setelah kedua data diolah maka dapat diuji tingkat akurasi untuk metode Decision Tree. Penelitian ini adalah untuk menentukan data yang paling akurat dengan sebagai acuan diagnosis pada penyakit diabetes. Pengujian tingkat akurasi dilakukan menggunakan confusion matrix dan kurva ROC/AUC (Area Under Cover). Hasil perhitungan data training menggunakan Algoritma Decision Tree (C4.5). Diketahui tingkat akurasi adalah 85,78%. Tabel Confussion Matrix disajikan pada tabel 2 dan gambar 5 merupakan grafik ROC/AUC (Area Under Cover) dari model Confussion Matrix yaitu 0.784.

Tabel 2. Akurasi Decision Tree

accuracy: 85.78%

	true 0.0	true 1.0	class precision
pred. 0.0	40548	6710	85.80%
pred. 1.0	14	4	22.22%
class recall	99.97%	0.06%	



Gambar 5. Kurva Confusion Matrix ROC Decision Tree

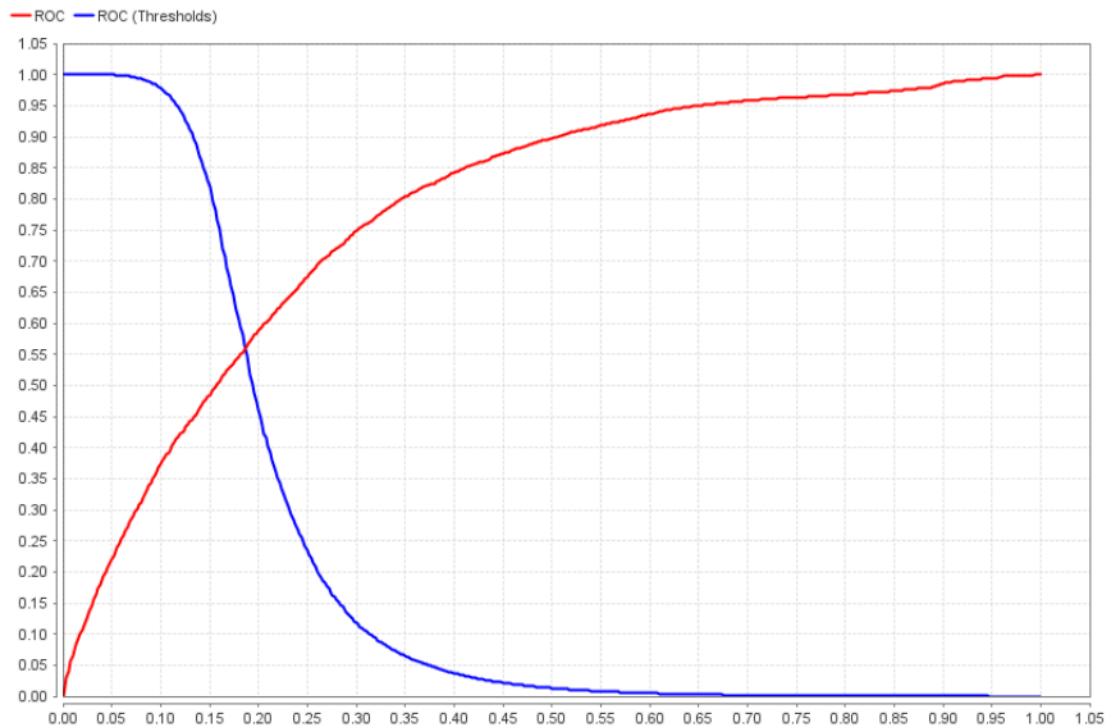
### b. Naïve Bayes

Pengujian selanjutnya dilakukan dengan menggunakan algoritma Naïve Bayes Diketahui tingkat akurasinya adalah 77,32%. Tabel Confussion Matrix disajikan pada tabel 3 dan gambar 6 merupakan grafik ROC/AUC (Area Under Cover) dari model Confussion Matrix yaitu 0.784.

Table 3 : Akurasi Naïve Bayes

accuracy: 77.32%

	true 0.0	true 1.0	class precision
pred. 0.0	32682	2844	91.99%
pred. 1.0	7880	3870	32.94%
class recall	80.57%	57.64%	



Gambar 6 : Kurva ROC Matrix Naïve Bayes

### Perbandingan Metode Decision Tree dan Naïve Bayes

	Decision Tree	Naïve Bayes
Accuracy	85,78 %	77,32 %



Lisensi  
 Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.



<b>AUC</b>	0,784	0,784
------------	-------	-------

Performance untuk pengakurasian AUC dapat diklasifikasikan lima kelompok yaitu:

1. 0.90 – 1.00 = Exellent Classification
2. 0.80 – 0.90 = Good Classification
3. 0.70 – 0.80 = Fair Classification
4. 0.60 – 0.70 = Poor Classification
5. 0.70 – 0.80 = Failure Classification

Berdasarkan klasifikasi dapat disimpulkan bahwa algoritma Decision Tree (C4.5) dan Naïve Bayes termasuk dalam Fair Classification (0.70 – 0.8)

### Kesimpulan dan Saran

Pada kesimpulan Dalam penelitian ini telah dilakukan pengujian pada data pasien penyakit diabetes menggunakan metode Decision Tree (C4.5) dan Naïve Bayes. Untuk mengetahui komparasi algoritma yang paling baik dalam menentukan penyakit liver. Untuk mengukur kinerja kedua metode tersebut digunakan metode pengujian Split Validation dan Cross Validation. Dapat disimpulkan bahwa metode Decision Tree (C4.5) dalam klasifikasinya menghasilkan akurasi 85,78%. dan nilai AUC 0.784 yang termasuk dalam Fair Classification. Naïve Bayes menghasilkan akurasi 77.32% dan nilai AUC 0.784. Dengan demikian dapat disimpulkan metode yang memberikan pecahan untuk permasalahan dalam mengidentifikasi penyakit diabetes ialah Decision Tree (C4.5).

### Daftar Pustaka

- [1] M. d. M. N. Arhami, *Data Mining Algoritma dan Implementasi*, Yogyakarta: CV Andi Offset (Penerbit ANDI, 2020).
- [2] d. E. T. L. Kusrini, *Algoritma Data Mining.*, Yogyakarta: Andi, 2009.
- [3] A. Kurniawan, “Sitem Pakar Untuk Mendiagnosa Penyakit Tuna Grahita Dengan Metode Naive Bayes,” in *Jurnal Ilmiah SINUS*, 2013.
- [4] E. Kusrini dan Taufiq, *Algoritma Data Mining*, Jogjakarta: CV. Andi Offset, 2009.
- [5] A. Kristanto, *Perancangan Sistem Informasi dan Aplikasinya*, Yogyakarta: Gava Media, 2018.
- [6] A. Kurniawan, “Sitem Pakar Untuk Mendiagnosa Penyakit Tuna Grahita Dengan Metode Naive Bayes,” in *Jurnal Ilmiah SINUS*, 2013.
- [7] Y. N. B. & S. F. Via, “Sistem Pendukung Keputusan Klasifikasi Tingkat Keganasan Kanker Payudara Dengan Metode Naive Bayes Classifier,” in *Scan, 2 (X)*, pp.64-65, 2015.
- [8] M. A. a. N. Fitrianiingsih, “Implementasi Algoritma Naive Bayes Classifier sebagai Sistem Rekomendasi Pembimbing Skripsi,” *InfoTekJar J. Nas. Inform. dan Teknol. Jar*, vol. vol. 5 no. 1, pp. pp. 44–50,, 2020.
- [9] F. L. a. H. Nugroho, “Klasifikasi Daun Herbal Menggunakan Metode Naïve Bayes Classifier Dan Knearest Neighbor,” *J. Simantec.*, Vols. vol. 5, no. 1,, p. pp. 9–16, 2015.

