

## DISKRITISASI *EQUAL-WIDTH INTERVAL* PADA NAÏVE BAYES (STUDI KASUS: KLASIFIKASI PASIEN TBC)

### *EQUAL-WIDTH INTERVAL DISCRETIZATION IN NAÏVE BAYES (CASE STUDY: CLASSIFICATION TBC PATIENTS)*

Hariyani<sup>a</sup>, Sugiyarto Surono<sup>b</sup>

Program Studi Matematika FAST UAD

Jl. Ringroad Selatan, Tamanan, Banguntapan, Bantul, Yogyakarta

[hariyani893@gmail.com](mailto:hariyani893@gmail.com), [sugiyartophd@gmail.com](mailto:sugiyartophd@gmail.com)

#### ABSTRAK

Klasifikasi Naive Bayes merupakan teknik untuk memprediksi probabilitas keanggotaan suatu kelas dengan menerapkan teorema Bayes. Klasifikasi Naive Bayes akan lebih baik jika menggunakan data yang berbentuk kategorik, sehingga dalam penelitian ini digunakan diskritisasi *equal-width interval* pada Naive Bayes untuk mengubah data yang berbentuk numerik menjadi kategorik. Tujuan dari penelitian ini adalah untuk menerapkan metode Naive Bayes dengan diskritisasi *equal-width interval* dalam mengklasifikasi pasien TBC di Puskesmas Sewon 1. Hasil penelitian ini menunjukkan akurasi sebesar 100% dengan perbandingan data *training* dan data *testing* sebesar 80%:20% dan 90%:10%, sehingga klasifikasi Naive Bayes dapat dikategorikan baik dalam mengklasifikasi pasien TBC.

**Kata Kunci :** diskritisasi, *equal-width interval*, klasifikasi, Naive Bayes, TBC

#### ABSTRACT

*The Naive Bayes classification is a technique for predicting the probability of class membership by applying the Bayes theorem. The Naive Bayes classification will be better if it uses categorical data, so in this study the Naive Bayes equal-width interval discretization is used to convert numeric data into categorical data. The purpose of this study was to apply the Naive Bayes method with equal-width interval discretization in classifying TB patients at the Sewon 1 Public Health Center. The results of this study showed an accuracy of 100% with a comparison of training data and testing data of 80%: 20% and 90%: 10%, so that the classification of Naive Bayes can be categorized as good in classifying TB patients.*

**Keywords :** discretization, *equal-width interval*, classification, Naive Bayes, TB

#### Pendahuluan

*Data mining* merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi – informasi yang nantinya dapat

digunakan (Saleh, 2015). *Data mining* mengacu pada penggalian atau penemuan pengetahuan yaitu proses menemukan pola yang menarik dalam kumpulan data besar untuk menemukan informasi pengambilan keputusan yang berguna. Kumpulan data

harus dibersihkan terlebih dahulu sebelum ditambang. Proses pembersihan ini menghilangkan kesalahan, memastikan konsistensi dan memperhitungkan nilai yang hilang. Kemudian, algoritma komputer digunakan untuk menambang data bersih untuk mencari pola yang tidak biasa. Selanjutnya, pola ditafsirkan untuk menghasilkan pengetahuan baru.

Klasifikasi merupakan metode *data mining* yang mengklasifikasikan data dalam kelas yang berbeda. Proses klasifikasi data terdiri dari dua tahap yaitu tahap pembelajaran dan tahap pengujian. Pada tahap pembelajaran, data *training* dianalisis dengan algoritma klasifikasi. Kemudian pada tahap pengujian, data *testing* digunakan untuk memperkirakan keakuratan aturan klasifikasi. Jika akurasi dianggap dapat diterima, maka aturan dapat diterapkan pada klasifikasi data baru (Han, Michelin & Jian, 2012). Salah satu teknik pengklasifikasian dalam *data mining* yaitu metode klasifikasi Naive Bayes.

Klasifikasi Naive Bayes adalah teknik prediksi berdasarkan probabilitas sederhana dan penerapan teorema Bayes (atau aturan Bayes) dengan asumsi independensi yang kuat (Salmi & Zuherman, 2019). Pengaruh nilai variabel pada kelas yang diberikan tidak tergantung pada nilai-nilai variabel yang lain. Terlepas dari kesederhanaannya, klasifikasi

Naive Bayes sering kali berhasil dengan sangat baik dan banyak digunakan karena sering mengungguli metode klasifikasi yang lain (Vembandasamy, dkk, 2015).

Berdasarkan penelitian Nur Nafi'iyah, Siti Mujilahwati (2018), input dataset dari algoritma Naive Bayes lebih baik dalam bentuk kategorik. Sehingga pada penelitian ini diterapkan metode diskritisasi *equal-width interval* untuk mengubah data yang berbentuk numerik menjadi bentuk kategorik. Contoh data dalam bentuk kategori yaitu besar, sedang, kecil, tinggi, dan lain-lain. Diskritisasi merupakan salah satu pendekatan populer untuk memetakan atribut numerik ke atribut kategorikal yang dapat mengurangi jumlah kesalahan klasifikasi NBC (Basica dkk, 2012).

Handayanna, dkk (2017) menggunakan metode Naive Bayes untuk memprediksi penyakit diabetes tipe 3 dengan optimasi parameter menggunakan algoritma genetika. Variabel yang digunakan antara lain berapa kali hamil, konsentrasi glukosa, tekanan darah, lipatan kulit, serum insulin, massa tubuh, diabetes tipe 3 silsilah fungsi, dan umur. Hasil yang diperoleh dari penelitian ini yaitu metode Naive Bayes mempunyai tingkat akurasi sebesar 72,00% dan metode Naive Bayes berbasis Algoritma

Genetika mempunyai tingkat akurasi sebesar 74,74%.

Asroni, dkk. (2018) menggunakan Algoritma Naive Bayes *Classifier* untuk memprediksi lama alumni UMY mendapatkan pekerjaan setelah menyelesaikan program studi S1. Variabel yang digunakan dalam penelitian ini yaitu jenis kelamin, fakultas, IPK, tahun lulus, dan tanggal mulai kerja. Hasil yang diperoleh yaitu alumni UMY diprediksi mendapatkan pekerjaan lebih cepat setelah menyelesaikan studi S1 dengan tingkat *accuracy* dari *performanceVector* yaitu 71,00%, *class precision* yaitu cepat 68,00%, lambat 75,00% dan untuk *class recall* yaitu cepat 83,02%, lambat 57,45%.

Menurut Kementerian Kesehatan RI, *Tuberculosis* (TBC) yang juga dikenal dengan TB adalah suatu penyakit menular yang disebabkan oleh kuman *Mycobacterium tuberculosis*. Terdapat beberapa jenis *Mycobacterium* antara lain: *M. tuberculosis*, *M. africanum*, *M. bovis*, *M. leprae* dan sebagainya. Gejala utama TBC berupa batuk berdahak lebih dari 2 minggu. Selain gejala batuk berdahak yang berlangsung lama, akan timbul gejala lain seperti sesak nafas, demam, berat badan turun, nafsu makan turun, badan lemas, nyeri dada, dan berkeringat di malam hari. Penularan penyakit *Tuberculosis* sangat

mudah yaitu melalui percikan ludah penderita TBC ketika batuk, berbicara, atau bersin. *Tuberculosis* biasanya menyerang paru-paru, tetapi juga bisa berdampak pada bagian tubuh lainnya. Seorang TB paru yang pemeriksaan kumannya positif, dalam satu tahun apabila tidak diobati akan menularkan kepada 15 orang lainnya. Oleh karena itu, untuk menyembuhkan penyakit TBC memerlukan pelayanan kesehatan yang cepat dan tepat. Sehingga, diperlukan suatu metode yang dapat mendiagnosa penyakit TBC secara dini.

Penemuan kasus TB BTA Positif di seluruh wilayah Kabupaten Bantul pada tahun 2019 menurut Dinas Kesehatan Kabupaten Bantul adalah sebesar 112,9 per 100.000 penduduk. Kecamatan dengan jumlah kasus TB BTA Positif terbanyak adalah Puskesmas Sewon 1. Tujuan dari penelitian ini adalah untuk menerapkan metode Naive Bayes dalam mengklasifikasi pasien TBC di Puskesmas Sewon 1.

### Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah metode Naive Bayes dengan penerapan diskritisasi *equal-width interval*. Objek penelitian yang digunakan yaitu data sekunder yang diperoleh dari data rekam medis pasien TBC di Puskesmas

Sewon 1. Adapun tahapan dalam menganalisis data yaitu sebagai berikut:

1. Mempelajari literatur yang berhubungan dengan TBC, *data mining*, metode Naïve Bayes, dan diskritisasi *equal-width interval*.
2. Mengambil data penelitian di Puskesmas Sewon 1.
3. Menginput data di Microsoft Excel.
4. Mengolah data pada python.
  - a. Melakukan *preprocessing* data.
  - b. Mengubah data yang berbentuk numerik menjadi bentuk kategorik dengan menggunakan diskritisasi *equal-width interval*.
  - c. Membagi data menjadi data *training* dan data *testing*.

- d. Memproses data *training* menggunakan metode Naïve Bayes sehingga diperoleh model klasifikasi Naïve Bayes.
- e. Menguji metode Naïve Bayes menggunakan *confusion matrix*.

### Hasil dan Pembahasan

Penelitian ini menggunakan data TBC sebanyak 52 data dengan variabelnya sebanyak 16 variabel. Variabel-variabelnya yaitu jenis kelamin ( $X_1$ ), umur ( $X_2$ ), batuk ( $X_3$ ), demam ( $X_4$ ), pilek ( $X_5$ ), sesak nafas ( $X_6$ ), nyeri dada ( $X_7$ ), diabetes mellitus ( $X_8$ ), nafsu makan turun ( $X_9$ ), berat badan turun ( $X_{10}$ ), dahak ( $X_{11}$ ), lemas ( $X_{12}$ ), nama ( $X_{13}$ ), HIV ( $X_{14}$ ), BTA ( $X_{15}$ ), dan diagnosa TBC ( $Y$ ). Berikut sampel data TBC yang digunakan:

**Tabel 1.** Sampel data TBC

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$	$X_{14}$	$X_{15}$	$Y$
P	37	Ya	Ya	Tdk	Ya	Tdk	Tdk	Tdk	Tdk	Ya	Tdk	H	NR	Negatif	Ya
L	11	Ya	Ya	Ya	Tdk	Tdk	Tdk	Tdk	Ya	Ya	Tdk	K	Tdk	-	Ya
L	3	Tdk	Ya	Tdk	Tdk	Tdk	Tdk	Tdk	Ya	Tdk	Tdk	N	Tdk	-	Tdk

Tahapan dalam mengolah data pada penelitian ini adalah sebagai berikut:

#### 1. *Preprocessing data*

Pada tahap ini akan dilakukan *preprocessing* data berupa seleksi data dan transformasi data.

- a. Seleksi data

Proses seleksi data yaitu proses mengambil data yang variabelnya sesuai untuk dianalisis. Pada tahap ini, variabel yang akan diambil yaitu jenis kelamin ( $X_1$ ), umur ( $X_2$ ), batuk ( $X_3$ ), demam ( $X_4$ ), pilek ( $X_5$ ), sesak nafas ( $X_6$ ), nyeri dada ( $X_7$ ), diabetes mellitus ( $X_8$ ), nafsu makan

turun ( $X_9$ ), berat badan turun ( $X_{10}$ ), dahak ( $X_{11}$ ), lemas ( $X_{12}$ ), dan diagnosa TBC ( $Y$ ). Sedangkan variabel nama ( $X_{13}$ ), HIV ( $X_{14}$ ), BTA ( $X_{15}$ ) akan dibuang karena tidak diperlukan dalam penelitian ini.

b. Transformasi data

Proses transformasi data yaitu proses mengubah data kebentuk yang sesuai untuk proses *data mining*. Transformasi yang dilakukan yaitu sebagai berikut:

1. Transformasi pada  $X_1$ :
  - a) P : 0
  - b) L : 1
2. Transformasi pada  $X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}$ :
  - a) Ya : 1
  - b) Tdk : 2
3. Transformasi pada  $X_2$  dilakukan dengan menerapkan proses diskritisasi.

**Proses Diskritisasi Data**

Proses diskritisasi data yaitu proses perubahan data dari data yang berbentuk numerik menjadi data yang berbentuk kategorik. Secara umum, jika diketahui atribut numerik  $X$  dan sampel acak  $\{x_i\}_{i=1}^n$  ukuran  $n$  yang diambil dari  $X$ , tugas diskritisasi yaitu membagi rentang nilai  $X$  menjadi  $k$

interval yang berurutan disebut juga *bins*, dengan mencari  $k-1$  nilai batas  $v_1, v_2, \dots, v_{k-1}$  yang menghasilkan interval

$$[x_{min}, v_1], (v_1, v_2], \dots, (v_{k-1}, x_{max}]$$

di mana nilai ekstrem dari rentang  $X$  diberikan sebagai berikut

$$x_{min} = \min\{x_i\}$$

$$x_{max} = \max\{x_i\}$$

$k$  interval yang dihasilkan dapat menjangkau seluruh rentang  $X$ . biasanya dipetakan ke nilai simbolik yang terdiri dari domain untuk atribut kategorikal baru dari  $X$  (Zaki & Wagner, 2014).

Penelitian ini menggunakan proses diskritisasi data *equal-width interval*.

***Equal-Width Interval***

Pendekatan binning yang paling sederhana yaitu membagi rentang  $X$  menjadi  $k$  interval yang sama lebar. Lebar interval yaitu rentang  $X$  dibagi dengan  $k$

$$w = \frac{x_{max} - x_{min}}{k} \quad (1)$$

Batas interval ke- $i$  diberikan sebagai berikut

$$v_i = x_{min} + iw, \text{ untuk } i = 1, 2, \dots, k - 1 \quad (2)$$

(Zaki & Wagner, 2014)

## 2. Pembagian data *training* dan data *testing*

Dalam penelitian ini, dari total keseluruhan data dilakukan variasi pembagian data *training* dan data *testing* sebagai berikut: 50%:50%, 60%:40%, 70%:30%, 80%:20%, dan 90%:10%.

## 3. Penerapan metode klasifikasi Naïve Bayes

Teorema Bayes dinamakan berdasarkan Thomas Bayes seorang ilmuwan Inggris yang mengemukakan teorema ini pertama kali.

**Teorema 1** (Papoulis, 1992)

Misalkan kejadian  $B_1, B_2, \dots, B_k$  merupakan suatu partisi dari  $S$  dengan  $P(B_i) \neq 0$  untuk  $i = 1, 2, \dots, k$ , maka untuk setiap kejadian  $A$  anggota  $S$

$$P(A) = \sum_{i=1}^k P(B_i \cap A) = \sum_{i=1}^k P(B_i)P(A|B_i) \quad (3)$$

**Teorema 2** (Walpole, 1995)

Misalkan kejadian  $B_1, B_2, \dots, B_k$  merupakan suatu partisi dari  $S$  dengan  $P(B_i) \neq 0$  untuk  $i = 1, 2, \dots, k$ . Misalkan  $A$  suatu kejadian sembarang dalam  $S$  dengan  $P(A) \neq 0$  maka

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^k P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^k P(B_i)P(A|B_i)} \quad (4)$$

Klasifikasi Naïve Bayes adalah klasifikasi statistik yang dapat memprediksi probabilitas keanggotaan suatu kelas, didasarkan pada teorema Bayes. Teorema

Bayes memiliki bentuk umum sebagai berikut:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (5)$$

Berikut akan dijabarkan rumus Bayes lebih lanjut sehingga menjadi klasifikasi Naïve Bayes. Misalkan  $Y$  merepresentasikan kelas dan  $X$  merepresentasikan variabel-variabel yang dibutuhkan untuk melakukan klasifikasi,  $X$  dapat ditulis sebagai berikut:

$$X = (x_1, x_2, \dots, x_n) \quad (6)$$

Jika ada  $m$  kelas, maka  $Y_i = Y_1, Y_2, \dots, Y_m$ . Dengan mensubstitusikan  $Y_i$  dan nilai  $X$  pada persamaan (5) maka akan diperoleh nilai probabilitas masing-masing kelasnya sebagai berikut:

$$P(Y_i|x_1, x_2, \dots, x_n) = \frac{P(Y_i)P(x_1, x_2, \dots, x_n|Y_i)}{P(x_1, x_2, \dots, x_n)} \quad (7)$$

Selanjutnya dilakukan penjabaran  $P(Y_i|x_1, x_2, \dots, x_n)$  menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned} P(Y_i|x_1, x_2, \dots, x_n) &= P(Y_i)P(x_1, x_2, \dots, x_n|Y_i) \\ &= P(Y_i)P(x_1|Y_i)P(x_2|Y_i) \dots \\ &\quad P(x_n|Y_i) \\ &= P(Y_i) \cdot \prod_{j=1}^n P(x_j|Y_i) \end{aligned} \quad (8)$$

Penyebut  $P(x_1, x_2, \dots, x_n)$  pada persamaan (7) diabaikan karena  $P(x_1, x_2, \dots, x_n)$  muncul di penyebut  $P(Y_i|x_1, x_2, \dots, x_n)$  untuk semua nilai  $j, j = 1, 2, \dots, n$ .  $P(x_1, x_2, \dots, x_n)$  konstan untuk setiap kelas. Hasil akhirnya yaitu posterior

sebanding dengan perkalian antara *prior* dengan *likelihood*, ditunjukkan pada persamaan berikut (Wiley, 2015):

$$P(Y_i|x_1, x_2, \dots, x_n) \propto P(Y_i) \cdot \prod_{j=1}^n P(x_j|Y_i)$$

$$i = 1, 2, \dots, m \quad (9)$$

Berikut tahapan klasifikasi Naïve Bayes:

- a. Menghitung nilai *prior* untuk setiap kelas.

$$P(Y_i) = \frac{n_i}{n}, \quad i = 1, 2, \dots, m$$

$n_i$  merupakan banyaknya kejadian  $I$  di  $Y$  dan  $n$  merupakan jumlah total kejadian.

- b. Menghitung nilai *likelihood* untuk setiap kelas.

$$P(X|Y_i) = \prod_{j=1}^n P(x_j|Y_i)$$

$$= P(x_1|Y_i) \cdot P(x_2|Y_i) \cdot \dots \cdot P(x_n|Y_i)$$

- c. Menghitung probabilitas akhir untuk setiap kelas dengan mengalikan nilai *prior* dan nilai *likelihood* yang sudah diperoleh sebelumnya.

$$P(Y_i|X) = P(Y_i) \cdot P(X|Y_i)$$

- d. Membandingkan hasil probabilitas akhir, lalu diambil nilai yang terbesar sebagai hasil klasifikasi.

*Maximum a posterior hypothesis* ( $h_{MAP}$ ) yaitu untuk menghasilkan nilai probabilitas posterior dari setiap kelas yang paling besar.

$$h_{MAP} = \arg \max P(Y_i|x_1, x_2, \dots, x_n)$$

$$= \arg \max \frac{P(Y_i)P(x_1, x_2, \dots, x_n|Y_i)}{P(x_1, x_2, \dots, x_n)}$$

$$\propto \arg \max P(Y_i) \cdot \prod_{j=1}^n P(x_j|Y_i)$$

#### 4. Pengujian metode klasifikasi Naïve Bayes

Pengujian metode klasifikasi Naïve Bayes pada penelitian ini menggunakan *confusion matrix*. *Confusion matrix* adalah sebuah table yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Berikut tabel pengukuran kinerja menggunakan *confusion matrix*:

**Tabel 2.** Pengukuran Kinerja menggunakan *Confusion Matrix*

		Klasifikasi	
		Positif	Negatif
Kelas	Positif	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
	Negatif	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

*True Positive* (TP) yaitu data positif yang diklasifikasikan benar sebagai data positif, *True Negative* (TN) yaitu data negatif yang diklasifikasikan benar sebagai data negatif, *False Positive* (FP) yaitu data negatif yang diklasifikasikan sebagai data positif, sedangkan *False Negative* (FN) yaitu data positif yang diklasifikasikan sebagai data negatif.

Berdasarkan *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN) diperoleh nilai akurasi, presisi, dan *recall*.

1. Akurasi yaitu perbandingan antara data yang terklasifikasi benar dengan data keseluruhan.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Presisi yaitu data positif yang diklasifikasikan benar dibagi dengan keseluruhan data yang diklasifikasikan benar.

$$Presisi = \frac{TP}{TP + FP}$$

3. *Recall* yaitu data positif yang diklasifikasikan benar oleh system.

$$Recall = \frac{TP}{TP + FN}$$

Berdasarkan Tabel 2 dapat dibuat *confusion matrix* sebagai berikut:

$$Confusion\ matrix = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}$$

Pengujian menggunakan python dilakukan dengan variasi data *training* dan data *testing* sebagai berikut 50%:50%, 60%:40%, 70%:30%, 80%:20%, dan 90%:10%. Dari pengujian ini, akan diketahui seberapa akurat klasifikasi Naïve Bayes dalam mengklasifikasikan data pasien TBC. Sebelum diuji menggunakan *confusion matrix*, data *testing* diklasifikasi terlebih dahulu. Hasil klasifikasi Naïve Bayes dari data *testing* dengan perbandingan data *training* dan data *testing* 70%:30% disajikan dalam Tabel 3.



**Tabel 3.** Hasil Klasifikasi Naïve Bayes

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$Y$	Klasifikasi
1	1	1	2	2	2	1	2	2	2	2	2	Ya	Ya
1	1	1	1	2	1	2	2	2	1	2	2	Ya	Ya
1	1	1	1	1	1	2	2	2	2	2	2	Ya	Ya
1	2	1	2	2	2	2	2	2	1	2	2	Ya	Ya
0	2	1	1	2	2	1	2	2	2	2	2	Ya	Ya
0	2	1	2	2	2	2	2	2	2	1	2	Ya	Ya
1	2	1	2	2	1	1	2	2	2	2	2	Ya	Ya
1	1	1	2	2	2	1	2	2	1	2	2	Ya	Ya
1	2	1	2	2	1	2	2	2	2	1	2	Ya	Ya
1	2	1	2	2	2	2	1	1	1	2	2	Ya	Ya
0	2	1	2	2	1	2	2	2	1	1	2	Ya	Ya
1	1	1	1	1	2	2	2	2	1	1	2	Ya	Ya
1	2	1	2	2	2	2	2	2	2	2	2	Tdk	Ya
0	1	1	2	2	1	1	2	2	2	2	2	Tdk	Ya
0	1	1	1	1	1	2	1	2	2	2	1	Tdk	Ya
0	1	1	2	2	1	2	1	2	2	2	2	Ya	Ya

Hasil pengujian Naïve Bayes menggunakan *confusion matrix* pada python dengan variasi data *training* dan data *testing*

sebagai berikut 50%:50%, 60%:40%, 70%:30%, 80%:20%, dan 90%:10% disajikan dalam Tabel 4.

**Tabel 4.** Hasil Pengujian Naïve Bayes

Variasi Data <i>Training</i> dan Data <i>Testing</i>	Hasil Pengujian	
50%:50%	Akurasi	0,7692
	Presisi	0,8300
	<i>Recall</i>	0,9100
60%:40%	Akurasi	0,8095
	Presisi	0,8100
	<i>Recall</i>	1,0000
70%:30%	Akurasi	0,8125
	Presisi	0,8125
	<i>Recall</i>	1,0000
80%:20%	Akurasi	1,0000
	Presisi	1,0000
	<i>Recall</i>	1,0000
90%:10%	Akurasi	1,0000
	Presisi	1,0000
	<i>Recall</i>	1,0000

Akurasi menunjukkan keakuratan model dalam mengklasifikasi, presisi menunjukkan keakuratan data yang dimiliki dengan hasil klasifikasi, sedangkan *recall* menunjukkan data positif yang diklasifikasikan benar oleh sistem.

### Kesimpulan

Dari pengujian Naïve Bayes menggunakan *confusion matrix* dengan variasi perbandingan data *training* dan data *testing* sebagai berikut 50%:50%, 60%:40%, 70%:30%, 80%:20%, dan 90%:10%, diperoleh akurasi terbesar yaitu sebesar 100%

pada perbandingan data *training* dan data *testing* 80%:20% dan 90%:10%. Sehingga dapat disimpulkan bahwa metode Naïve Bayes dengan penerapan diskritisasi *equal-width interval* dapat mengklasifikasi pasien TBC dengan baik.

### Pustaka

- Asroni, Nadiyah M. A., dan Slamet R. 2018. Perkiraan Masa Tunggu Alumni Mendapatkan Pekerjaan menggunakan Metode Prediksi Data Mining dengan Algoritma Naïve Bayes Classifier. *Semesta Teknika*. Vol. 21, No. 2, 189-197.
- Basica, Warih Maharani, Intan Nurma Yulita. 2012. Analisis dan Implementasi

- Genetic Algoritma Fuzzy Discretization untuk Naïve Bayes Classifier [Tugas Akhir]. Bandung(ID): Universitas Telkom.
- Han, Jiawei, Michelin K., and Jian P. 2012. *Data Mining Concepts and Techniques*. USA: Elsevier.
- Handayanna, Frisma, dkk. 2017. Prediksi Penyakit Diabetes menggunakan Naïve Bayes dengan Optimasi Parameter menggunakan Algoritma Genetika. *Konferensi Nasional Ilmu Sosial dan Teknologi (KNiST)*, pp.71-76.
- Nafi'iyah, Nur dan Siti Mujilawati. 2018. Analisis Algoritma Backpropagation dan Naïve Bayes dalam Identifikasi Jenis Kelamin Manusia Berdasarkan Foto Panoramik Gigi. *Seminar Nasional Sistem Informasi 2018*.
- Papoulis, Athanasios. 1992. *Probabilitas, Variabel random, dan Proses Stokastik*. Diterjemahkan oleh Dr. Subanar. Yogyakarta: Gajah Mada University Press.
- Saleh, Alfa. 2015. Implementasi Metode Klasifikasi Naïve Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal*. Vol. 2, No. 3.
- Salmi, Nafizatus and Zuherman Rustam. 2019. Naïve Bayes Classifier Models for Predicting the Colon Cancer. *9<sup>th</sup> Annual Basic Science International Conference 2019*.
- Vembandasamy, K., dkk. 2015. Heart Diseases Detection Using Naïve Bayes Algorithm. *International Journal of Innovative Science, Engineering and Technology*. Vol. 2 Issue 9.
- Walpole, Ronald E. dan Raymond H. Myers. 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan*. Diterjemahkan oleh R. K. Sembiring. Bandung: Penerbit ITB.
- Wiley. 2015. *Data Science and Big Data Analytics*. Canada: John Wiley and Sons, Inc
- Zaki, Mohammed J. and Wagner M. Jr. 2014. *Data Mining and Analysis: Fundamental Concepts and Algorithms*. New York: Cambridge University Press.

