

PENERAPAN METODE NAÏVE BAYES DALAM MENGKLASIFIKASI TINGKAT OBESITAS PADA PRIA

Juan Veron Wie¹, Muhammad Siddik^{2*}

^{1,2} Prodi Teknik Informatika, Fakultas Ilmu Komputer, Institut Bisnis dan Teknologi Pelita Indonesia
e-mail: ¹juanveronwie@student.pelitaIndonesia.ac.id, ²siddik@lecturer.pelitaIndonesia.ac.id

Abstract

Obesity is a disease caused by abnormal excess fat in adipose tissue. The results of basic health research in 2018 said that the population aged over 18 years who were obese had increased from 14.8% to 21.8%. Obesity can also increase the risk of critical illnesses such as heart disease and stroke, therefore, it is quite important to predict whether someone is obese or not so that obesity can be treated early. Here the data taken is only anthropometric data measurements on male subjects, this is because the composition of body fat in men has a higher level of muscle mass in the body, men are also more likely to have adipose tissue in the abdominal area. In this study, secondary data was used which was taken from kaggle.com. This data has 15 variables and 252 data with 2 classifications, namely (Obesity or Non Obesity). From the results of the classification test at a ratio of 70% training and 30% testing using naïve Bayes, the results obtained were Accuracy 88%, Precision (Obesity) 73%, Precision (Non Obesity) 98%, Recall (Obesity) 96% with Recall (Non Obesity) 85% and the F1-Score result is 0.830188679. These results show an average scale above 80% so that it can be said that the Naïve Bayes Classifier (NBC) algorithm is able to classify obesity in men well.

Keywords: Classification, Naïve Bayes, Obesity

Abstrak

Obesitas merupakan salah satu penyakit yang disebabkan kelebihan lemak yang tidak normal pada jaringan adiposa. Hasil riset kesehatan dasar pada tahun 2018 mengatakan penduduk yang berumur lebih dari 18 tahun yang mengalami obesitas meningkat dari 14.8% menjadi 21.8%. Penyakit obesitas juga dapat meningkatkan risiko penyakit kritis seperti penyakit jantung dan stroke, oleh karena itu, cukup penting dilakukan prediksi terhadap seseorang terkena obesitas atau tidak agar obesitas dapat ditangani lebih awal. Disini data yang diambil hanyalah pengukuran data antropometri pada subjek pria hal ini dikarenakan komposisi lemak tubuh pada pria memiliki tingkat massa otot pada tubuh yang lebih tinggi, pria juga lebih cenderung untuk memiliki jaringan adiposa pada area abdomen. Pada penelitian ini digunakan data sekunder yang diambil dari kaggle.com. Data ini memiliki 15 variabel dan 252 data dengan 2 klasifikasi yaitu (Obesitas atau Non Obesitas). Dari hasil pengujian klasifikasi pada rasio 70% training dan 30% testing menggunakan naïve bayes didapatkan hasil *Accuracy* 88%, *Precision (Obesitas)* 73%, sedangkan *Precision (Non Obesitas)* 98%, *Recall (Obesitas)* 96% dengan *Recall (Non Obesitas)* 85% dan hasil *F1-Score* sebesar 0.830188679. Hasil tersebut menunjukkan skala rata-rata diatas 80% sehingga dapat dikatakan bahwa algoritma Naïve Bayes Classifier (NBC) mampu mengklasifikasi obesitas pada pria dengan baik.

Kata kunci: Klasifikasi, Naïve Bayes, Obesitas

1. PENDAHULUAN

Obesitas adalah sebuah kondisi kronis yang diakibatkan karena konsumsi kalori berlebihan, obesitas dapat ditandai dengan adanya penumpukan lemak dalam tubuh yang sangat tinggi. Obesitas terjadi karena asupan

kalori yang lebih banyak dibanding aktivitas membakar kalori atau dalam kata lain kalori yang masuk tidak sebanding dengan kalori yang dikeluarkan. (dr. Tjin Willy, 2018). Di Indonesia, menurut Riset Kesehatan Dasar

(Riskesdas) tahun 2007 melaporkan bahwa 11,7% orang dewasa berumur ≥ 18 tahun mengalami obesitas dan berdasarkan Riset Kesehatan Dasar (Riskesdas) tahun 2010 meningkat 11,65%. (Listiyana et al., 2013)

Gizi lebih dan obesitas merupakan faktor risiko terjadinya penyakit jantung koroner disamping faktor risiko lainnya, seperti hipertensi, diabetes melitus, merokok, stres, dan kurang olahraga. (Sandjaja, 2014). Menurut Seidell dan Halberstadt dalam jurnal (Septiyanti & Seniwati, 2020) menjelaskan bahwa obesitas juga berhubungan dengan peningkatan inflamasi dan metabolisme tubuh yang abnormal, sehingga meningkatkan risiko resistensi insulin, diabetes melitus tipe 2, stroke, dan penyakit kardiovaskular. (Patonah et al., 2019). Obesitas dapat ditentukan melalui pemeriksaan secara antropometri. Pengukuran antropometri untuk usia dewasa sekarang ini menggunakan perhitungan Indeks Massa Tubuh (IMT), dengan perbandingan (rasio) berat badan / tinggi badan² yang sering digunakan untuk menilai status gizi orang dewasa. (Nyoman et al., 2017)

Dataset antropometri 252 pria didapat dari (Garth Fisher, 2021) yang meliputi beberapa pengukuran antropometri pada pria dilihat dari beberapa variabel seperti densitas tubuh, persen lemak tubuh, usia, berat badan (pon), tinggi badan (inci), lingkar leher (cm), lingkar dada (cm), lingkar perut (cm), lingkar pinggang (cm), lingkar paha (cm), lingkar mata kaki (cm), lingkar lutut (cm), lingkar mata kaki (cm), lingkar bicep (cm), lingkar lengah bawah (cm), lingkar pergelangan tangan (cm). Dimana setelah dilihat dari hasil data tersebut terdapat 65 pria dengan persen lemak tubuh yang sangat tinggi yaitu melebihi 25% yang sudah diklasifikasikan sangat tinggi atau sudah mencapai tingkat obesitas, walaupun mayoritas hasil dari klasifikasi IMT 65 pria ini menunjukkan status gizi pada ambang normal dan pre-obesitas namun ternyata tingkat persen lemak tubuh pada sejumlah pria ini sudah diambang obesitas. Disini data yang diambil hanyalah pengukuran data antropometri pada subjek pria hal ini dikarenakan komposisi lemak tubuh pada pria dan pada wanita berbeda, dengan pria memiliki tingkat massa otot pada tubuh yang lebih tinggi dibanding wanita, pria juga lebih cenderung untuk memiliki jaringan adiposa pada area abdomen, sedangkan pada wanita

lebih banyak dilihat pada bagian paha dan panggul. (Tina et al., 2021).

Dari permasalahan yang didapat diatas peneliti ingin memprediksi obesitas menggunakan algoritma *Naïve Bayes* dengan beberapa kriteria yang didapat dari pengukuran antropometri pada pria beserta perhitungan Indeks Massa Tubuh (IMT) dan digunakanlah perhitungan persen lemak tubuh berdasarkan perhitungan lingkar perut dan lingkar leher sebagai kriteria utama, dengan mengimplementasikan algoritma *Naïve Bayes* diharapkan dapat melakukan perhitungan persentase lemak tubuh sehingga bisa mengklasifikasi dan memprediksi tingkat obesitas pada pria. Kemudian hasil dari klasifikasi tersebut akan menghasilkan dua buah kategori yaitu kategori obesitas dan tidak obesitas berdasarkan variabel yang telah ditentukan sebelumnya.

2. METODE PENELITIAN

Pada penelitian ini penulis menggunakan metode klasifikasi algoritma *Naïve Bayes*. Menurut Jiawei dalam jurnal (Miranda & Julisar, 2018) menjelaskan bahwa keunggulan menggunakan metode *naïve bayes classifier* dibandingkan metode lainnya adalah merupakan metode pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan dari suatu kelas, selain itu terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database yang besar. Ciri utama dari *Naïve Bayes Classifier* (NBC) adalah asumsi yang sangat kuat (*naïf*) akan independensi dari masing-masing kondisi atau kejadian. (Syarli & Muin, 2016).

Dimana metode ini akan diterapkan dengan dataset yang sudah tersedia dan mengolah dataset berdasarkan kebutuhan untuk menentukan obesitas pada pria. sehingga dihasilkan bentuk data pada gambar di bawah ini.

WG	H	N	C	AB	HP	T	K	AK	B	F	W	BMI	BF	BFC
154.3	67.8	36.2	93.1	85.2	94.5	59	37.3	21.9	32	27.4	17.1	23.6	19.05	NO
173.3	72.3	38.5	93.6	83	98.7	58.7	37.3	23.4	30.5	28.9	18.2	23.3	13.50	NO
154	66.3	34	95.8	87.9	99.2	59.6	38.9	24	28.8	25.2	16.6	24.6	23.29	NO
184.8	72.3	37.4	101.8	86.4	101.2	60.1	37.3	22.8	32.4	29.4	18.2	24.9	17.10	NO
184.3	71.3	34.4	97.3	100	101.9	63.2	42.2	24	32.2	27.7	17.7	25.5	28.42	O
210.3	74.8	39	104.5	94.4	107.8	66	42	25.6	35.7	30.6	18.8	26.4	20.65	NO
181	69.8	36.4	105.1	90.7	100.3	58.4	38.3	22.9	31.9	27.8	17.7	26.1	22.00	NO
176	72.5	37.8	99.6	88.5	97.1	60	39.4	23.2	30.5	29	18.8	23.5	18.29	NO
191	74	38.1	100.9	82.5	99.9	62.9	38.3	23.8	35.9	31.1	18.2	24.5	12.71	NO
198.3	73.5	42.1	99.6	88.6	104.1	63.1	41.7	25	35.6	30	19.2	25.8	14.64	NO
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
186.3	74.5	38.5	101.5	83.6	98.2	59.7	39.7	25.2	32.8	29.4	18.5	23.6	13.09	NO
216	76	39.4	103.6	90.9	107.7	66.2	39.2	25.9	37.2	30.2	19	26.3	17.44	NO
180.5	69.5	38.4	102	91.6	103.9	63.4	38.3	21.5	32.5	28.6	17.7	26.3	21.37	NO
205.3	71.3	39.4	104.1	101.8	108.6	66	41.5	23.7	36.9	31.6	18.8	28.4	26.55	O
187.8	69.5	40.5	101.3	96.4	100.1	69	39	23.1	36.1	30.5	18.2	27.3	23.22	NO
162.8	66	36.4	99.1	92.8	99.2	63.1	38.7	21.7	31.1	26.4	16.9	26.3	25.13	O
195.8	71	38.9	101.9	96.4	105.2	64.8	40.8	23.1	36.2	30.8	17.3	27.3	23.63	NO
209.3	71	42.1	107.6	97.5	107	66.9	40	24.4	38.2	31.6	19.3	29.2	22.24	NO
183.8	67.8	38	106.8	89.6	102.4	64.2	38.7	22.9	37.2	30.5	18.5	28.1	20.98	NO
211.8	73.5	40	106.2	100.5	109	65.8	40.6	24	37.1	30.1	18.2	27.6	24.47	NO

Gambar 1. Potongan data selection diambil dari kaggle.com 252 data.

Ini 15 variabel yang meliputi pengukuran antropometri pada pria.

WG : Berat Badan (kg)

H : Tinggi Badan (m)

N : Lingkar Leher (inci)

C : Lingkar Dada (inci)

AB : Lingkar Perut (inci)

HP : Lingkar Pinggang (cm)

T : Lingkar Paha (cm)

K : Lingkar Lutut (cm)

AK : Lingkar Mata Kaki (cm)

B : Lingkar Bisep (cm)

F : Lingkar Lengan Bawah (cm)

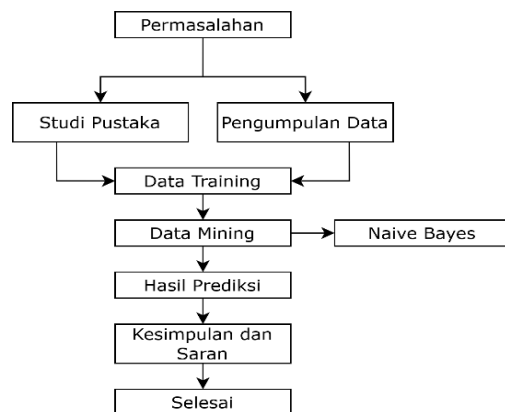
W : Lingkar Pergelangan Tangan (cm)

BMI : Indeks Massa Tubuh

BF : Persen Lemak Tubuh (%)

BFC : Klasifikasi (Obesitas atau Non Obesitas)

Kerangka penelitian dapat dilihat pada gambar di bawah ini



Gambar 2. Kerangka Penelitian

Identifikasi permasalahan perlu dilakukan untuk menggambarkan masalah yang terjadi, sehingga didapatkan rumusan masalah, tujuan dan manfaat penelitian. Kemudian mencari teori-teori untuk melengkapi sumber kajian yang bersumber dari jurnal, buku, dan internet. Mencari sumber data yang diperlukan pada penelitian ini data yang digunakan adalah data pengukuran antropometri yang didapat dari Kaggle Machine Learning and Data Science Community dengan alamat web <https://www.kaggle.com/>, dari 252 data perlu dilakukan proses cleaning sehingga dihasilkan 251 record data dengan 1 record data outlier, penghapusan record ini harus dilakukan karena dapat mengganggu hasil akhir klasifikasi obesitas. Pembagian data training dan data testing digunakan rasio 70% training dan 30% testing (175 record data training dan 76 record data testing). Data yang didapat akan diolah ke format yang dibutuhkan, pengelompokan dan penentuan variabel dari data awal. Data training yang digunakan dalam melatih algoritma *Naïve Bayes* perhitungan ini menghasilkan nilai keluaran yang diinginkan. Pengujian yang dilakukan dan menampilkan hasil prediksi, sehingga nantinya dapat disimpulkan berdasarkan dari hasil yang didapat.

2.1 Indeks Massa Tubuh

Menurut (Apriyanti et al., 2018) Indeks Massa Tubuh (IMT), tidak lain angka yang menunjukkan apakah seseorang terlalu berat untuk tinggi mereka. IMT juga dapat digunakan untuk menentukan seberapa besar resiko dapat terkena resiko penyakit tertentu yang terkait karena berat badan. Klasifikasi berat badan pada orang dewasa berdasarkan IMT menurut WHO 2000 dapat dilihat pada tabel berikut:

Tabel 1. Klasifikasi Berat Badan Pada Orang Dewasa

Kategori	BMI (Body Mass Index)
Underweight	< 18,5 kg/m ²
Batas normal	18,5 – 24,9 kg/m ²
Overweight	≥ 25 kg/m ²
Pre-obese	25,0 – 29,9 kg/m ²
Obese I	30,0 – 39,9 kg/m ²
Obese II	35,0 – 39,9 kg/m ²

Obese III $\geq 40,0 \text{ kg/m}^2$

IMT dihitung dengan menggunakan rumus sebagai berikut:

$$IMT = \frac{Bobot(kg)}{[Tinggi(m)]^2} \quad (1)$$

2.2 Persen Lemak Tubuh

Berdasarkan dari pengukuran antropometri tubuh untuk mendapatkan pengukuran lingkaran perut, lingkaran leher dan tinggi badan. Persen lemak tubuh untuk pria dapat diestimasi dengan kalkulasi menggunakan rumus dari US Navy Body Fat Formula yaitu :

$$BF (\%) = 86.010 \times \log_{10}(LP - LL) - 70.041 \times \log_{10}(TB) + 36.76 \quad (2)$$

2.3 Naïve Bayes Classifier

Naïve Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan (Siddik et al., 2019)

Teorema Bayes memiliki bentuk umum sebagai berikut (Muslehatin et al., 2017):

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (3)$$

Keterangan:

X = Data dengan class yang belum diketahui
H = Hipotesis data X merupakan suatu class spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X (posteriori prob.)

$P(H)$ = Probabilitas hipotesis H (prior prob.)

$P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut

$P(X)$ = Probabilitas dari X

Langkah-langkah yang ada dalam metode Naïve Bayes Classifier menurut (Guntur et al., 2018) adalah sebagai berikut :

1. Baca data training
2. Hitung jumlah dan probabilitas
 - a. Jika terdapat data numerik, maka rumus yang digunakan untuk menghitung nilai rata-rata hitung (mean) dapat dilihat sebagai berikut:

$$\mu = \sum_{i=1}^n X_i \quad (4)$$

atau

$$\mu = \frac{X_1 + X_2 + X_3 + X_4 + \dots + X_n}{n} \quad (5)$$

Dimana :

μ : rata-rata hitung (mean)

X_i : nilai sample ke -i

n : jumlah sampel

- b. Jika tidak, hitung nilai probabilitas tiap kategori yang sama, dengan jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut untuk menemukan nilai probabilitistik.

3. Menghitung nilai simpangan baku (standar deviasi) dapat dilihat dibawah ini:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \quad (6)$$

Dimana :

σ : standar deviasi

x_i : nilai x ke -i

μ : rata-rata hitung

n : jumlah sampel

4. Menghitung data uji, jika terdapat data numerik cari nilai distribusi gaussian masing-masing fitur. Jika tidak langsung ke langkah selanjutnya. Persamaan untuk menghitung nilai gaussian Naïve bayes dapat dilihat dibawah ini :

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (7)$$

Dimana :

P : Peluang

X_i : Atribut ke -i

x_i : Nilai Atribut ke -i

Y : Kelas yang dicari

y_i : Sub-kelas yang dicari

μ : mean, menyatakan rata-rata dari seluruh atribut

σ : Standar Deviasi, menyatakan varian dari seluruh atribut

5. Probabilitas akhir tiap kelas, memasukkan dan menghitung semua data probabilitas ke dalam satu kelas yang sama.
6. Probabilitas akhir, mengalikan probabilitas tiap kelas dengan probabilitas akhir tiap kelas. Berikut Persamaan untuk perhitungan berdasarkan probabilitas :

$$P(X_1, \dots, X_k | C) = P(X_1 | C) \times \dots \times P(X_k | C) \quad (8)$$

Jika atribut ke-i bersifat diskret, maka $P(X_i | C)$ diestimasi sebagai frekuensi

relatif dari sampel yang memiliki nilai x_i sebagai atribut ke i dalam kelas C . Namun, jika atribut ke- i bersifat kontinu, maka $P(X_i | C)$ diestimasi dengan fungsi densitas Gauss.

2.4 Pengujian Naïve Bayes

Penelitian ini melakukan evaluasi dengan menggunakan beberapa metrik yang diperlukan untuk menggambarkan performa dari sebuah model atau algoritma secara spesifik menggunakan beberapa persamaan. Persamaan ini memiliki bentuk sebagai berikut (Saputro & Sari, 2020):

1. Akurasi

Adalah total keseluruhan seberapa sering model benar mengklasifikasi. Formula akurasi yaitu menggunakan persamaan dibawah ini.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

Keterangan:

TP = True Positive (jumlah data positif yang terklasifikasi dengan benar oleh sistem)

TN = True Negative (jumlah data negatif yang terklasifikasi dengan benar oleh sistem)

FP = False Positive (jumlah data positif namun terklasifikasi salah oleh sistem)

FN = False Negative (jumlah data positif namun terklasifikasi salah oleh sistem)

2. Precision

Ketika model memprediksi positif atau negatif, seberapa sering prediksi itu benar. Formula Precision yaitu menggunakan persamaan dibawah ini.

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

3. Recall

Ketika kelas aktualnya positif, seberapa sering model memprediksi positif. Formula recall yaitu menggunakan persamaan dibawah ini.

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

4. F1-Score

Merupakan rata-rata harmonik dari Precision dan Recall. Formula F1-Score yaitu menggunakan persamaan dibawah ini.

$$F1 - Score = 2 * \frac{Precision * recall}{Precision + recall} \quad (11)$$

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan penelitian yang tergambar dari perhitungan pengujian Naïve Bayes sebagai berikut :

3.1 Data Training

Data training adalah data latih yang digunakan dalam melatih algoritma Naïve Bayes Classifier dalam mengolah perhitungan dalam menghasilkan nilai keluaran yang diinginkan terdiri dari data testing dengan rasio 70% training dan 30% testing. Data training yang digunakan adalah data yang sudah dikumpulkan sebelumnya, yakni 175 record data training dan 76 record data testing.

a. Mencari Nilai Mean dan Standar Deviasi

Mencari nilai mean dan standar deviasi variabel obesitas dan non obesitas berdasarkan klasifikasi pada data training. Untuk menghitung nilai mean dengan klasifikasi obesitas, dihitung jumlah obesitas yang terdapat pada tabel data training yaitu sebanyak 30 data. Nilai dari 30 data tadi dijumlahkan, lalu dibagi dengan jumlah data obesitas. Kemudian menggunakan rumus untuk mencari mean untuk atribut BERAT dengan hasil di klasifikasi obesitas sebesar :

$$\mu = \frac{83.56 + 93.08 + 112.13 + 86.96 + \dots + 102.83}{30}$$

$$\mu (obesitas) = 94.54$$

Kemudian setelah nilai μ pada masing-masing atribut telah didapatkan maka selanjutnya kita mencari nilai standar deviasi. Untuk menghitung nilai standar deviasi klasifikasi obesitas, jumlah data terdapat 30 data untuk klasifikasi obesitas, lalu dipangkat 2, setelah itu dijumlahkan dan dibagi dengan jumlah data dikurangi 1, lalu diakarkan. Dengan menggunakan rumus standar deviasi maka didapatkan :

$$\sigma = \sqrt{\frac{(83.56 - 95.54)^2 + (93.08 - 95.54)^2 + \dots + (102.83 - 95.54)^2}{30 - 1}}$$

$$\sigma (obesitas) = 16.18$$

b. Nilai Probabilitas Fitur Setiap Kelas

Setelah menghitung nilai mean dan standar deviasi untuk fitur dengan data angka, maka selanjutnya mencari probabilitas setiap fitur pada kelas. Jumlah data yang digunakan adalah sebanyak 175 data. Data obesitas ada 30 dan data non obesitas ada 138. Sehingga untuk menghitung nilai probabilitasnya adalah sebagai berikut :

$$P(\text{Obesitas}) = \frac{30}{175} = 0.171428571$$

$$P(\text{Non Obesitas}) = \frac{145}{175} = 0.828571429$$

Setelah mencari nilai probabilitas tiap fitur, maka didapat tabel untuk nilai mean, standar deviasi dan probabilitas, bisa dilihat pada tabel berikut.

Tabel 2. Tabel Probabilitas

Probabilitas Iya			
Nilai	Mean	Std	Std^2
Berat	81.74	11.7	136.8
Tinggi	1.78	0.06	0
Leher	15.55	1.06	1.11
Dada	108.89	7.79	60.66
Perut	41.42	3.73	13.95
Pinggang	106.82	9.1	82.83
Paha	64.47	5.59	31.27
Lutut	40.33	2.31	5.34
Mata Kaki	23.74	1.56	2.43
Bisep	34.32	2.84	8.06
Lengan	29.52	2.09	4.35
Pergelangan	18.68	0.97	0.94
IMT	25.91	4.27	18.23
Lemak Tubuh	28.82	3.8	14.43

Probabilitas Tidak			
Nilai	Mean	Std	Std^2
Berat	80.24	0.07	0.75
Tinggi	1.79	0.86	35.32
Leher	14.73	5.94	7.89
Dada	97.58	2.81	26.42
Perut	34.59	5.14	19.29
Pinggang	97.82	4.39	3.82
Paha	58.06	1.95	3.37
Lutut	37.89	1.84	7.83
Mata Kaki	22.96	2.8	3.59
Bisep	31.6	1.9	0.73

Lengan	28.42	0.86	20.07
Pergelangan	18.04	4.48	18.02
IMT	25.13	4.25	0
Lemak Tubuh	18.71	0	0
Klasifikasi	O		0.21
	NO		0.79

c. Nilai Distribusi Gaussian

Untuk menghitung nilai distribusi gaussian, menggunakan rumus sebagai berikut:

$$P(\text{BERAT} = 68.48 | \text{Obesitas}) = \frac{1}{\sqrt{2\pi}11.8573} 2.7183^{-\frac{(68.48-81.7)^2}{2*140.6}}$$

$$P(\text{BERAT} = 68.48 | \text{Obesitas}) = 0.0620334$$

Nilai diatas adalah nilai distribusi gaussian untuk atribut BERAT hasil obesitas yang didapat dengan memasukkan nilai data uji BERAT, nilai μ dan nilai σ BERAT. Dihitung juga nilai distribusi gaussian pada atribut atribut lainnya dengan langkah yang sama.

d. Probabilitas Akhir Setiap Kelas

Menghitung probabilitas akhir untuk setiap kelas dengan memasukan semua data nilai distribusi gaussian yang ada ke dalam satu kelas yang sama, seperti dibawah ini.

$$P(\text{BERAT} | O) = P(\text{BERAT} = 68.48 | O) \times P(H1 = 67 | O) \times \dots \times P(BF = 18.57 | O)$$

$$P(\text{BERAT} | O) = 1.83111879094468E - 17$$

Nilai diatas adalah probabilitas akhir untuk kelas dengan hasil obesitas. Untuk mencari probabilitas kelas dengan hasil non obesitas, langkah-langkah nya sama seperti menghitung probabilitas akhir kelas dengan hasil obesitas

e. Probabilitas Akhir

Probabilitas akhir didapat melalui perhitungan nilai probabilitas akhir kelas. Perhitungan probabilitas akhir adalah sebagai berikut.

$$P(\text{Obesitas} | \text{BERAT}) = 0.211428571428571 \times 1.83111879094468E - 17$$

$$P(\text{Obesitas} | \text{BERAT}) = 3.87150830085446E - 18$$

Setelah mendapatkan probabilitas akhir, dilakukan normalisasi untuk memperoleh nilai sama dengan 1, dengan cara sebagai berikut:

$$P(\text{Obesitas}) = \frac{3.87150830085446E - 18}{3.87150830085446E - 18 + 0.999999997748201} = 2.25179888570916E - 09$$

Hasil diatas adalah hasil probabilitas akhir untuk kelas obesitas. Kemudian data uji yang diolah sebanyak 76 data. Berikut data uji yang diujikan.

Tabel 3. Hasil Klasifikasi

No	P Normalisasi		Klasifikasi	Data Riil
	O	NO		
1	9E-09	1E+00	NO	NO
2	1E+00	4E-15	O	O
3	1E-01	9E-01	NO	NO
4	1E+00	9E-11	O	NO
5	1E+00	6E-07	O	O
6	4.-15	1E+00	NO	NO
7	3E-09	1E+00	NO	NO
8	7E-08	1E+00	NO	NO
9	2E-07	1E+00	NO	NO
10	9E-09	1E+00	NO	NO
11	1E+00	1E-11	O	O
12	1E+00	2E-03	O	NO
13	1E+00	4E-03	O	NO
14	9E-01	5E-02	O	O
15	5E-11	1E+00	NO	NO
16	1E+00	1E-14	O	O
17	1E-01	9E-01	NO	NO
18	1E+00	6E-07	O	O
19	1E-08	1E+00	NO	NO
20	9E-02	9E-01	NO	NO
21	2E-07	1E+00	NO	NO
22	5E-07	1E+00	NO	NO
23	8E-11	1E+00	NO	NO
24	4E-04	1E+00	NO	NO
25	5E-05	1E+00	NO	NO
26	2E-09	1E+00	NO	NO
27	1E+00	2E-03	O	O
28	1E-05	1E+00	NO	NO
29	1E+00	6E-12	O	O
30	9E-01	1E-01	O	NO
31	5E-04	1E+00	NO	NO
32	9E-01	1E-01	O	NO
33	4E-09	1E+00	NO	NO
34	4E-09	1E+00	NO	NO

35	4E-11	1E+00	NO	NO
36	1E+00	3E-06	O	O
37	7E-07	1E+00	NO	NO
38	8E-01	2E-01	O	NO
39	2E-08	1E+00	NO	NO
40	1E+00	4E-18	O	O
41	2E-10	1E+00	NO	NO
42	3E-10	1E+00	NO	NO
43	1E+00	9E-03	O	NO
44	7E-09	1E+00	NO	NO
45	2E-05	1E+00	NO	NO
46	1E+00	6E-13	O	O
47	5E-07	1E+00	NO	NO
48	4E-10	1E+00	NO	NO
49	4E-02	1E+00	NO	NO
50	2E-12	1E+00	NO	NO
51	2E-04	1E+00	NO	NO
52	7E-01	3E-01	O	NO
53	9E-05	1E+00	NO	NO
54	3E-04	1E+00	NO	NO
55	4E-08	1E+00	NO	NO
56	6E-03	1E+00	NO	NO
57	1E-06	1E+00	NO	NO
58	4E-06	1E+00	NO	O
59	3E-06	1E+00	NO	NO
60	3E-04	1E+00	NO	NO
61	1E-01	9E-01	NO	NO
62	1E+00	4E-12	O	O
63	9E-09	1E+00	NO	NO
64	1E+00	1E-03	O	O
65	1E-11	1E+00	NO	NO
66	1E+00	3E-16	O	O
67	1E+00	2E-12	O	O
68	1E+00	2E-13	O	O
69	1E+00	8E-06	O	O
70	8E-09	1E+00	NO	NO
71	1E+00	1E-10	O	O
72	4E-10	1E+00	NO	NO
73	1E+00	1E-05	O	O
74	1E+00	9E-05	O	O
75	9E-01	6E-02	O	O
76	1E+00	6E-09	O	O

Berdasarkan hasil klasifikasi yang terdapat pada tabel 3 menjelaskan bahwa penentuan obesitas maupun non obesitas dapat dilihat pada bagian P Normalisasi yaitu nilai probabilitas akhir untuk (O) Obesitas dan

(NO) Non Obesitas. Jika nilai probabilitas akhir obesitas (O) lebih besar dari probabilitas akhir non obesitas (NO) maka hasil yang diperoleh berada pada kelas klasifikasi obesitas (O) dan jika nilai probabilitas akhir kelas non obesitas (NO) lebih besar maka diperoleh hasil kelas klasifikasi non obesitas (NO).

Kemudian dilakukan pengujian *Naïve Bayes* dengan cara menghitung akurasi, precision, recall dan f1-scor pengujian model algoritma *naïve bayes classifier* pada tabel berikut :

Tabel 4. Hasil Pengujian

80-20	True O	True NO
Pred O	9	7
Pred NO	1	33
Accuracy	84%	
Precision	56%	97%
Recall	90%	83%
f1 Score	0.692307692	
70-30	True O	True NO
Pred O	22	8
Pred NO	1	45
Accuracy	88%	
Precision	73%	98%
Recall	96%	85%
f1 Score	0.830188679	
60-40	True O	True NO
Pred O	26	6
Pred NO	6	62
Accuracy	88%	
Precision	81%	91%
Recall	81%	91%
f1 Score	0.8125	

Dari hasil pengujian klasifikasi pada rasio 70% training dan 30% testing menggunakan *naïve bayes* pada Ms.Excel didapatkan hasil *Accuracy* 88%, *Precision* (Obesitas) 73%, sedangkan *Precision* (Non Obesitas) 98%, *Recall* (Obesitas) 96% dengan *Recall* (Non Obesitas) 85% dan hasil *F1-Score* sebesar 0.830188679. Hasil tersebut menunjukkan skala rata-rata diatas 80% sehingga dapat dikatakan bahwa algoritma *Naïve Bayes Classifier* (NBC) mampu

mengklasifikasi obesitas pada pria dengan baik.

Dari hasil pengujian yang terlihat pada tabel 4 dilakukan 3 kali pengujian sehingga dapat disimpulkan bahwa semakin besar rasio data testing yang digunakan maka ada juga peningkatan jumlah kesalahan data yang diprediksi oleh sistem namun tingkat kesalahan yang diprediksi sistem tidak terlalu besar karena pada percobaan ketiga dengan rasio 60% training dan 40% testing tingkat akurasi masih diatas 80% dengan adanya total kesalahan prediksi sebesar 12 pada data obesitas dan non obesitas

4. SIMPULAN

Pengujian menggunakan 60% data *training* dan 40% data *testing* menghasilkan akurasi lebih tinggi sebesar 84.15%. Disini nilai akurasi yang didapat lebih tinggi dikarenakan pada *split* data 60:40 adanya lebih banyak data testing untuk pria Obesitas (O) yang digunakan sehingga jumlah data obesitas yang ada lebih besar jumlahnya dan karena dataset yang digunakan lebih dominan terhadap pria yang Non Obesitas (NO) maka nilai akurasi yang didapat bisa lebih besar.

Penerapan algoritma *Naïve Bayes* dalam memprediksi tingkat obesitas pada pria dapat diterapkan dengan menghasilkan tingkat akurasi yang memuaskan sebesar 84.15% pada rasio *split data* 60:40.

5. DAFTAR PUSTAKA

- Apriyanti, L., Rahayu, M. I., & Suseno, N. M. (2018). Aplikasi mobile pengelolaan kalori harian untuk penderita obesitas. *Jurnal Teknologi Informasi Dan Komunikasi*, 7(2), 16–23.
- dr. Tjin Willy. (2018). *Obesitas - Gejala, penyebab dan mengobati - Alodokter*.
- Garth Fisher, D. A. (2021). *Body Fat Prediction Dataset*. 2021.
- Guntur, M., Santony, J., & Yuhandri, Y. (2018). Prediksi Harga Emas dengan Menggunakan Metode *Naïve Bayes* dalam Investasi untuk Meminimalisasi Resiko. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 2(1), 354–360. <https://doi.org/10.29207/resti.v2i1.276>
- Listiyana, A. D., Mardiana, M., & Prameswari, G. N. (2013). Obesitas sentral dan kadar kolesterol darah total. *Jurnal Kesehatan*

- Masyarakat*, 9(1), 37–43.
- Miranda, E., & Julisar. (2018). DATA MINING DENGAN METODE KLASIFIKASI NAÏVE BAYES UNTUK MENGLASIFIKASIKAN PELANGGAN Eka Miranda , Julisar Program Sistem Informasi , Program Studi Sistem Informasi , Universitas Bina Nusantara. *Infotech*, 4(9), 6–12.
- Muslehatin, W., Ibnu, M., & Mustakim. (2017). Penerapan Naïve Bayes Classification untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau. *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI)*, 250–256.
- Nyoman, N., Yuliani, S., Subagio, H. W., & Murbawani, E. A. (2017). Korelasi Lingkar Leher Dengan Persentase Lemak Tubuh Pada Obesitas. *Journal of Nutrition and Health*, 5(3), 1–8. <https://doi.org/10.14710/jnh.5.3.2017.138-145>
- Patonah, P., Marlioni, L., & Mulyani, Y. (2019). Edukasi Pola Hidup Sehat Kepada Masyarakat Di Kelurahan Manjahlega Kota Bandung Dalam Menanggulangi Obesitas Sebagai Faktor Resiko Penyakit Kardiovaskular. *Amaliah: Jurnal Pengabdian Kepada Masyarakat*, 3(2), 354–361. <https://doi.org/10.32696/ajpkm.v3i2.290>
- Sandjaja, S. (2014). Prevalensi Gizi Lebih Dan Obesitas Penduduk Dewasa Di Indonesia. *Gizi Indonesia*, 28(2), 1–7. <https://doi.org/10.36457/gizindo.v28i2.21>
- Saputro, I. W., & Sari, B. W. (2020). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*, 6(1), 1. <https://doi.org/10.24076/citec.2019v6i1.178>
- Septiyanti, S., & Seniwati, S. (2020). Obesity and Central Obesity in Indonesian Urban Communities. *Jurnal Ilmiah Kesehatan (JIKA)*, 2(3), 118–127. <https://doi.org/10.36590/jika.v2i3.74>
- Siddik, M., Desnelita, Y., & Gustientiedina. (2019). Penerapan Naïve Bayes untuk Memprediksi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademis. *Jurnal Infomedia*, 2(4), 89–93.
- Syarli, S., & Muin, A. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). *Jurnal Ilmiah Ilmu Komputer*, 2(1), 22–26.
- Tina, A. R., Kurniawan, L. B., Bahrin, U., Pascasarjana, P., Biomedik, I., Pascasarjana, S., Hasanuddin, U., Klinik, D. P., Kedokteran, F., Hasanuddin, U., Perintis, J., No, K., & Selatan, S. (2021). *Analisis Hubungan Berbagai Indeks Obesitas dengan Interleukin-6 pada Subjek Obesitas dan Non Obesitas Sentral Kadar Association Between Obesity Indices and Interleukin-6 Levels in Subjects with Central and Non-Central Obesity*. 3(2), 104–113.