

Klasifikasi Data Antropometri Individu Menggunakan Algoritma *Naïve Bayes Classifier*

Johnson Sihombing

Sistem Informasi, STMIK Ganesha Bandung, john97.sihombing@gmail.com

Keywords:

*Decision Support System,
Text classification,
Naïve Bayes Classifier,
Python*

ABSTRACT

With the development of advances in computer technology today, most companies and organizations need a decision support system based on information systems, where the information is generally stored in the form of documents / text that is not structured. In this regard, a system for text management that is integrated with the decision support system is needed. One of them is the use of text data classification for anthropometric case studies of several samples. Anthropometry is a measurement of a person's body dimensions. The object of research is gender, first name, and height of a person. The research aims to determine the ratio of the number and height probability level of the number of men and women based on the input into an application using the Naïve Bayes Classifier method. The implementation design uses the Python programming language. The results showed that the height classification data frequency of women was more than the height classification data for men. And the number of height probability of a woman's body is greater than the number of height probability of a man's body.

Kata Kunci

*Sistem Pendukung Keputusan,
Klasifikasi Teks,
Naïve Bayes Classifier,
Python*

ABSTRAK

Dengan berkembangnya kemajuan teknologi komputer saat ini, sebagian besar perusahaan maupun organisasi membutuhkan suatu sistem pendukung keputusan yang berbasis sistem informasi, dimana informasi tersebut pada umumnya tersimpan dalam bentuk dokumen/teks yang tidak terstruktur. Sehubungan hal tersebut, dibutuhkan suatu sistem untuk pengelolaan teks yang terintegrasi dengan sistem pendukung keputusan. Salah satunya adalah penggunaan klasifikasi data berbentuk teks untuk studi kasus antropometri dari beberapa sampel. Antropometri adalah pengukuran dimensi tubuh seseorang. Yang menjadi objek penelitian adalah jenis kelamin, nama awal, dan tinggi seseorang. Penelitian ini bertujuan untuk menentukan perbandingan jumlah serta tingkat probabilitas tinggi tubuh jumlah pria dan wanita berdasarkan inputan kedalam satu aplikasi dengan metode Naïve Bayes Classifier. Perancangan implementasi menggunakan bahasa pemrograman Python. Hasil penelitian menunjukkan bahwa frekuensi kemunculan data klasifikasi tinggi tubuh wanita lebih banyak daripada data klasifikasi tinggi tubuh pria. Serta jumlah probabilitas tinggi tubuh wanita lebih besar dari jumlah probabilitas tinggi tubuh pria

Korespondensi Penulis:

Johnson Sihombing,
STMIK Ganesha Bandung,
Jl. Raya Militer KM 2 Kp. Cilamega Desa Depok Kec Darangdan, Purwakarta.
Telepon: +6282111165381
Email: john97.sihombing@gmail.com

1. PENDAHULUAN

1.1 Latar Belakang

Teknologi informasi merupakan salah satu bidang yang saat ini berkembang dengan pesat di segala bidang kehidupan. Data/informasi yang melimpah akan dihasilkan oleh teknologi informasi, mulai dari sektor industri,

ekonomi, serta sektor aplikasi lainnya. Penerapan teknologi informasi dalam kehidupan rumah tangga juga dapat menghasilkan data yang berlimpah mengenai antropometri (ukuran tubuh seseorang).

Pada jaman serba digital dewasa ini, masih terdapat beberapa bidang yang masih menggunakan cara-cara manual untuk mendapatkan data atau informasi, salah satunya adalah bidang antropometri (ukuran tubuh). Dan cara untuk mengukur antropometri seseorang adalah dengan kursi antropometri.

Antropometri merupakan suatu ilmu yang cenderung meneliti ke arah yang berkaitan dalam desain ruang kerja, fasilitas kerja, serta desain produk-produk yang sesuai dengan ukuran tubuh manusia. Tujuan dilakukannya hal tersebut adalah terciptanya suatu kondisi yang nyaman, aman, dan sehat bagi manusia serta menciptakan kondisi kerja yang efisien dengan hasil efektif untuk mencapai keadaan yang ergonomis.

Contoh dampak negatif apabila kita mendesain ruang kerja (interaksi antara operator dengan kursi, meja dan seperangkat komputer) tidak menggunakan antropometri manusia adalah desain *workstation* tersebut apabila tidak ergonomis dapat menyebabkan penyakit *Occupational Cervicobrachial Syndrome* (OCS), dan kelelahan mata akibat dari lamanya menatap layar monitor atau akibat dari posisi monitor yang tidak sesuai dengan *user* [1]. Dengan tersedianya data antropometri tubuh manusia Indonesia, maka dapat dideskripsikan ukuran yang presisi dan akurat sesuai dengan ukuran dimensi tubuh manusia Indonesia.

Di Indonesia sendiri, antropometri utamanya digunakan untuk meneliti status gizi pada anak. Hal ini dijelaskan pada Keputusan Menteri Kesehatan No 1995/MENKES/S/2010. Di mana dalam keputusan tersebut pengukuran antropometri mengikuti standar yang telah ditetapkan oleh WHO pada tahun 2005. Pada keputusan tersebut juga dijelaskan mengenai bagaimana tolak ukur yang digunakan dalam mengukur status gizi anak. Beberapa *variable* yang dicari dalam hal tersebut adalah: umur, panjang badan, tinggi badan, serta berat badan.

Pengukuran antropometri banyak digunakan dalam berbagai kegiatan, diantaranya dalam perekrutan calon pegawai negeri sipil, perekrutan tentara dan sebagai alat bantu dalam praktikum ergonomi [2].

Berdasarkan pendapat-pendapat dan keputusan tersebut di atas, penulis tertarik dan berminat untuk melakukan penelitian terhadap klasifikasi data antropometri tubuh manusia dengan menggunakan algoritma *Naïve Bayes*. Adapun proses implementasi dilakukan dengan bahasa pemrograman Python.

1.2 Tujuan Penelitian

Adapun tujuan yang dilaksanakan penulis pada penelitian ini adalah sebagai berikut:

1. Menghasilkan informasi untuk mendukung proses pengambilan keputusan dalam menentukan jumlah jenis kelamin dan tinggi seseorang.
2. Membuat aplikasi untuk menghitung probabilitas dan frekuensi terhadap komponen-komponen yang terkait dengan pengambilan keputusan.

2. METODE PENELITIAN

2.1 Antropometri

A. Pengertian Antropometri

Berdasarkan pendapat Aritonang Irianton (2013): "antropometri adalah suatu cara penilaian dari pengaruh faktor genetik dan faktor lingkungan terhadap status energi dan protein seseorang". Antropometri memiliki kelebihan-kelebihan: disamping prosedurnya yang cukup sederhana, juga tingkat keamanan yang tinggi, dan dapat dilakukan dalam jumlah sampel yang besar dan pada umumnya tidak membutuhkan tenaga ahli untuk menangani masalah tersebut [3].

Masih menurut pendapat Aritonang Irianton (2013): jenis pengukuran antropometri yang dilakukan terhadap seseorang adalah berat badan (BB), tinggi badan (TB), lingkaran lengan atas (LILA), lingkaran kepala (LK), tebal lemak dibawah kulit (TL), serta tinggi lutut [3].

Beberapa aspek yang menjadikan alat tersebut menjadi favorit para konsumen adalah: tidak terlalu mahal, fleksibilitas tinggi, berdaya tahan lama, serta mudah dipesan dan dibuat. Sedangkan kebalikannya adalah pendeteksian status gizi seseorang dapat berlangsung dalam waktu yang relatif lama. Tingkat akurasi dan keabsahan dari pengukuran antropometri akan berkurang jika terjadi kesalahan dalam pengukuran antropometri seseorang [4].

B. Ukuran Antropometri

Berikut ini adalah jenis-jenis ukuran yang dilakukan untuk antropometri:

a. Berat Badan (BB)

Berat badan menggambarkan tentang massa tubuh. Dalam keadaan normal, BB berkembang mengikuti perkembangan umur (balita). Sedangkan saat dalam keadaan tidak normal, BB berkembang lebih cepat atau lambat. Berdasarkan sifat tersebut, maka indikator BB/U hanya dapat menggambarkan status gizi saat ini. Prosedur penimbangan BB yaitu (1) dilakukan sebaiknya pagi hari setelah buang air atau keadaan perut kosong supaya hasil akurat, (2) meletakkan timbangan di tempat yang datar, (3) sebelum dilakukan penimbangan sebaiknya timbangan dikalibrasi terlebih dahulu, (4) klien diminta melepas alas kaki, aksesoris yang digunakan dan menggunakan pakaian seminimal mungkin, (5) klien naik ke timbangan dengan posisi menghadap kedepan, pandangan lurus, tangan disamping kanan kiri dan posisi rileks serta tidak banyak gerakan, (6) catat hasil pengukuran [3].

b. Tinggi Badan (TB).

Tinggi badan merupakan gambaran pertumbuhan. Dalam keadaan normal, TB tumbuh bersama dengan penambahan umur. Pengaruh kekurangan gizi terhadap TB akan tampak pada kekurangan yang sangat lama. Berdasarkan hal tersebut indeks TB/U dapat menggambarkan keadaan masa lalu (Aritonang, 2013). Prosedur pengukuran TB yaitu (1) memasang mikrotoa pada dinding yang rata dan tegak lurus pada lantai, (2) mikrotoa digeser keatas hingga melebihi tinggi anak yang akan diukur, (3) klien berdiri tegak lurus rapat ke dinding, (5) posisi kepala, bahu belakang, pantat dan tumit rapat ke dinding, pandangan lurus ke depan, (6) membaca angka pada mikrotoa dengan pandangan mata sejajar dengan angka yang ditunjuk pada garis mikrotoa [3].

2.2 Naïve Bayes Classifier

A. Pengertian Metode Naïve Bayes Classifier

Naïve Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yg sangat kuat akan independensi dari masing-masing kondisi/ kejadian.

Menurut Olson Delen (2008) menjelaskan Naïve Bayes untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Sedangkan jumlah frekuensi yang terdapat pada tabel keputusan merupakan faktor penentu dari probabilitas terhadap perkiraan akhir [5].

Metode *Naïve Bayes Classifier* bekerja sangat baik dibanding dengan model classifier lainnya. Hal ini dibuktikan oleh Xhemali, Hinde Stone dalam jurnalnya “Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages” mengatakan bahwa “*Naïve Bayes Classifier* memiliki tingkat akurasi yg lebih baik dibanding model classifier lainnya”. Disisi lain, dengan jumlah data pelatihan (training data) yg tidak terlalu besar, maka cukup mudah untuk menentukan perkiraan parameter yang dibutuhkan dalam klasifikasi datanya. Karena yang diasumsikan sebagai variabel independent, maka hanya varians dari suatu variabel dalam sebuah kelas yg dibutuhkan unt menentukan klasifikasi, bukan keseluruhan dari matriks kovarians. Adapun kegunaan dari metode *Naïve Bayes Classifier* adalah:

- Mengklasifikasikan dokumen teks seperti teks berita ataupun teks akademis.
- Sebagai metode machine learning yang menggunakan probabilitas.
- Untuk membuat diagnosis medis secara otomatis.
- Mendeteksi atau menyaring spam.

Metode *Naïve Bayes Classifier* memiliki kelebihan-kelebihan, antara lain [6]:

- Bisa dipakai untuk data kuantitatif maupun kualitatif.
- Tidak memerlukan jumlah data yang banyak.
- Tidak perlu melakukan data training yang banyak.
- Jika ada nilai yang hilang, maka bisa diabaikan dalam perhitungan.
- Proses perhitungan yang cepat dan efisien.
- Mudah dipahami.
- Mudah dibuat.
- Pengklasifikasian dokumen bisa dipersonalisasi, disesuaikan dengan kebutuhan setiap orang.
- Coding* yang cukup sederhana jika diimplementasikan kedalam bahasa pemrograman.
- Bisa digunakan untuk klasifikasi masalah biner ataupun *multiclass*.

Metode *Naïve Bayes Classifier* juga memiliki kekurangan-kekurangan, yaitu:

- Apabila probabilitas kondisionalnya bernilai nol, maka probabilitas prediksi juga akan bernilai nol.
- Asumsi bahwa masing-masing variabel independen yang akan berdampak terhadap berkurangnya akurasi, karena biasanya ada korelasi antara variabel yang satu dengan variabel yang lain.
- Keakuratan tidak bisa hanya diukur dengan menggunakan satu probabilitas saja, tapi butuh bukti-bukti lain untuk membuktikannya.
- Untuk membuat keputusan, diperlukan pengetahuan awal atau pengetahuan mengenai masa sebelumnya. Keberhasilan sangat bergantung pada pengetahuan awal tersebut.

B. Teorema Naïve Bayes Classifier

Menurut Bustami (2013): pada teorema Bayes, bila terdapat dua kejadian yang terpisah (misalkan X dan H), maka Teorema Bayes dirumuskan sebagai berikut [7] :

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

Keterangan:

X = Class data yang dimasukkan;

- H = Data hipotesis;
- P(H|X) = Probabilitas hipotesis H yang mengacu pada kondisi X;
- P(H) = Probabilitas hipotesis H;
- P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H.
- P(X) = Probabilitas X

Teorema Bayes sering pula dikembangkan mengingat berlakunya hukum probabilitas total, menjadi seperti berikut:

$$P(H|X) = \frac{P(X|H)}{\sum_{i=1}^n P(H_i|X)} \cdot P(H)$$

Keterangan:

- i = 1,2,3, ... , n adalah jumlah data hipotesis.
- dimana H1 U H2 U H3 ... U Hn = S.
- S = Probabilitas total H

Untuk menentukan jenis kelas yang cocok dalam pengklasifikasian, maka rumus diatas disesuaikan sebagai berikut [7] :

$$P(C|F1, ..., Fn) = \frac{P(F1, ..., Fn|C)}{P(F1, ..., Fn)} \cdot P(H)$$

Keterangan:

- C = representasi dari kelas.
- F1 ...Fn = karakteristik petunjuk yang dibutuhkan dalam proses klasifikasi data.

Formula diatas tersebut mendeskripsikan tentang peluang munculnya kelas **C** (sebelum masuknya *sampel*, disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik *sampel* pada kelas **C** (disebut juga *likelihood*), kemudian dibagi dengan peluang kemunculan karakteristik-karakteristik *sampel* secara global (disebut juga *evidence*). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut :

$$\text{Posterior} = \frac{\text{Prior} \times \text{likelihood}}{\text{Evidence}}$$

Dengan asumsi bahwa masing-masing petunjuk (**F1, F2, ..., Fn**) saling bebas (*independen*) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut :

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i \cdot F_j)}{P(F_j)} = P(F_i)$$

Untuk $i \neq j$, sehingga $P(F_i|C, F_j) = P(F_i|C)$ atau dapat ditulis dalam notasi :

$$P(c|F_1, F_2, F_3, \dots, F_n) = P(C) \prod_{i=1}^n P(F_i|C)$$

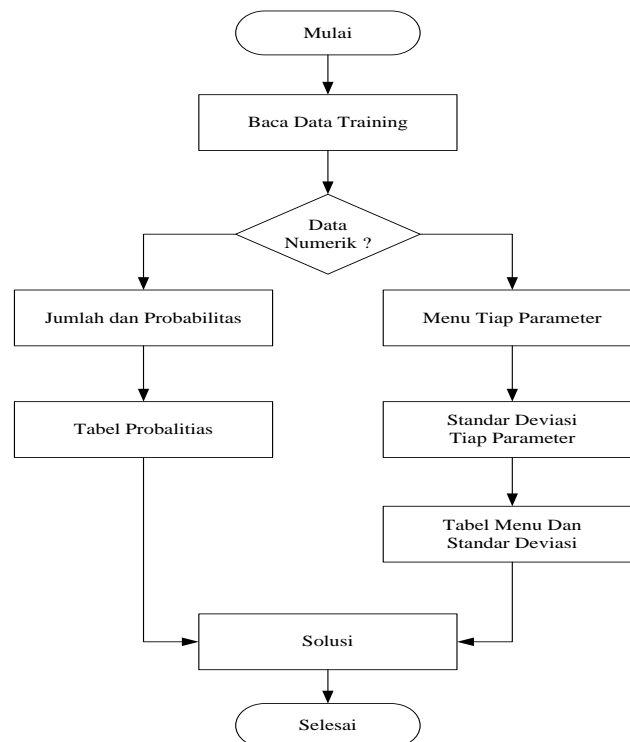
Untuk klasifikasi dengan data kontinyu digunakan rumus **Densitas Gauss** :

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

Keterangan:

- 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*, yang menyatakan rata-rata dari seluruh atribut ; σ = *Deviasi Standar*, yang menyatakan varian dari seluruh atribut

Menurut Saleh (2015), diagram alir dari metode *Naive Bayes* dapat dibuat seperti pada gambar dibawah ini [8] :



Gambar 1. Diagram Alir *Naive Bayes*

2.3 Laplace Correction

Manning (2009) berpendapat bahwa *Laplace Correction (Laplacian Estimator)* adalah suatu cara untuk menangani nilai probabilitas 0 (nol). Dengan melakukan perhitungan datanya yang ditambah 1 (satu), maka dapat menghindari nilai probabilitas 0 (nol) pada proses klasifikasi. *Laplace Correction* ini dapat dinyatakan dalam persamaan seperti pada persamaan berikut ini [9] :

$$p_i = \frac{m_i + 1}{n + k}$$

dimana p_i adalah probabilitas dari atribut m_i , m_i adalah jumlah sampel dalam kelas dari atribut m_i , k adalah jumlah kelas dari atribut m_i dan n adalah jumlah sampel.

Contoh: diasumsikan ada *class Pembelian BBM = sedang* di suatu training set, memiliki 8 sampel, ada 0 sampel dengan *Pengeluaran/Bulan = sedikit*, 4 sampel dengan *Pengeluaran/Bulan = sedang*, dan 4 sampel dengan *Pengeluaran/Bulan = banyak*.

Probabilitas dari kejadian ini tanpa Laplacian Correction adalah:

$$P(\text{Pengeluaran/Bulan} = \text{sedikit} | \text{Pembelian BBM} = \text{sedang}) = 0,$$

$$P(\text{Pengeluaran/Bulan} = \text{sedang} | \text{Pembelian BBM} = \text{sedang}) = 0.500 \text{ (dari 4/8), dan}$$

$$P(\text{Pengeluaran/Bulan} = \text{banyak} | \text{Pembelian BBM} = \text{sedang}) = 0.500 \text{ (dari 4/8).}$$

Dengan menggunakan Laplacian Correction dari tiga kasus diatas, diasumsikan ada satu sampel lagi untuk masing-masing nilai Pembelian BBM = sedikit. Dengan cara ini, didapatkan probabilitas sebagai berikut:

$$P(\text{Pengeluaran/Bulan} = \text{sedikit} | \text{Pembelian BBM} = \text{sedang}) = 0.090 \text{ (dari 1/11),}$$

$$P(\text{Pengeluaran/Bulan} = \text{sedang} | \text{Pembelian BBM} = \text{sedang}) = 0.454 \text{ (dari 5/11), dan}$$

$$P(\text{Pengeluaran/Bulan} = \text{banyak} | \text{Pembelian BBM} = \text{sedang}) = 0.454 \text{ (dari 5/11).}$$

2.4 Python Software

Python pada dasarnya memiliki *style* yang serupa dengan *pseudocode*, perbedaannya hanyalah Python dapat dijalankan di komputer dan menampilkan hasil (Hall, 2005, p 10). Bahasa pemrograman Python mulai dikembangkan pada tanggal 20 Februari 1991 oleh Van Rossum. Python adalah sebuah bahasa pemrograman dinamis yang sering digunakan dalam pengembangan aplikasi pada berbagai domain. Bahasa Python mudah dimengerti (serupa dengan

bahasa Inggris), dan tidak perlu proses kompilasi, maka dari itu kita dapat menjalankan *script* Python begitu kita menyimpannya tanpa memerlukan langkah lanjutan.

2.5 Hardware Pengujian

Perangkat keras yang digunakan penulis untuk implementasi aplikasi ini adalah sebagai berikut :

- OS : Windows 10 Pro
- Processor : Intel® Core™ i3-5005U
- CPU : @ 2.00 GHz 2.00 GHz
- RAM : 4.00 GB
- System type : 64-bit Operating System, x64-based operation
- Display Adapter : Intel(R) HD Graphics 5500 NVIDIA GeForce 930M

2.6 Makalah Penulis Lain

Untuk melengkapi penelitian, penulis juga membaca referensi-referens dari jurnal terkait seperti yang terlihat di bawah ini:

Tabel 1. Riset Penulis Lain

No	Penulis	Tahun	Publikasi	Judul
1	Karelana Indriani, Qonita Tanjung	2017	Publikasi Jurnal Penelitian Teknik Informatika Universitas Prima Indonesia (UNPRI) Medan Vol. 1, No. 2, Oktober 2018 e-ISSN: 2621-234x	Sistem Pendukung Keputusan Kelayakan Kredit Motor Menggunakan Metode Naive Bayes Pada NSC Finance Cikampek
Kesimpulan: perusahaan mudah dan cepat untuk input data pendaftar, hasil analisa metode Naive Bayes dapat diperoleh dengan cepat, tingkat akurasi sebesar 99% dan <i>error</i> 1% yang dibantu dengan aplikasi pendukung Rapidminer				
2	Harry Pribadi Fitriani, Ikhwan Ruslianto, Rahmi Hidayati	2018	Jurnal Coding, Sistem Komputer Untan Volume 06, No. 02 (2018), hal 13-24 ISSN: 2338-493X	Implementasi Metode Naive Bayes Classifier Untuk Aplikasi Filtering Email Spam Dengan Lemmatization Berbasis Web
Kesimpulan: Waktu proses <i>filtering file email</i> dengan ukuran yang lebih kecil akan lebih cepat dari ukuran <i>file email</i> yang lebih besar. Presentase keberhasilan filtering dari 113 <i>file email</i> adalah sebesar 90,83%				
3	Mohamat Dodi Trisetiyo, Jati Sasongko Wibowo	2019	Prosiding SENDU_U 2019 ISBN: 978-979-3649-99-3	Klasifikasi Surat Menggunakan Metode Naive Bayes Pada Sistem Informasi Manajemen Surat
Kesimpulan: pada uji pertama, nilai akurasi rendah 33% karena hasil konversi file .pdf tidak sesuai. Uji kedua menghasilkan akurasi 88%. <i>Input</i> data manual menghasilkan tingkat akurasi yang cukup tinggi.				
4	Dina Lovita Sari, Feri Candra	2019	Jom FTEKNIK Volume 6 Edisi 1 Janari s/d Juni 2019	Aplikasi Persediaan Barang Bangunan Cerdas Menggunakan Metode Naive Bayes Berbasis Web
Kesimpulan: Aplikasi layak digunakan dengan tingkat akurasi 98,14%. Aplikasi mampu mengelola data-data terkait, transaksi dan pembuatan laporan.				

3. HASIL DAN ANALISIS

Setelah dilakukan *coding* terhadap studi kasus ini, maka dilakukan pengujian terhadap applikasi yang dibuat, yang terdiri dari beberapa tahapan:

3.1 Data Training

Langkah pertama yang dilakukan adalah proses baca data training. Pada kasus ini ada **100** data *training*/sampel. Adapun data *training* yang digunakan adalah data individu yang terdiri atas 50 pria dan 50 wanita, seperti berikut ini :

Tabel 2. Data Training

No	Nama	Tinggi (cm)	Berat (Kg)	Jenis Kelamin
1	Randy Carter	184	73.0	Pria
2	Stephanie Smith	149	52	Wanita
3	Cynthia Watson	174	63	Wanita
4	Jessie Morgan	175	67.0	Pria
5	Katherine Carter	183	81	Wanita
6	David Reed	187	60.0	Pria
7	Stephen Jones	192	96.0	Pria

No	Nama	Tinggi (cm)	Berat (Kg)	Jenis Kelamin
8	Jerry Allen	204	91.0	Pria
9	Billy Wright	180	66.0	Pria
10	Earl Green	184	52.0	Pria
11	Todd Bell	174	53.0	Pria
12	Martin White	177	91.0	Pria
13	Elizabeth Ross	138	37	Wanita
14	Kenneth Cox	200	82.0	Pria
15	Frank Perry	193	79.0	Pria
16	Wayne Sanders	189	79.0	Pria
17	Carol King	149	59	Wanita
18	Henry Smith	188	53.0	Pria
19	Alan Foster	187	81.0	Pria
20	Paul Henderson	187	99.0	Pria
21	Jack Lewis	190	81.0	Pria
22	Christina Foster	161	48	Wanita
23	Beverly Long	179	75	Wanita
24	Bobby Parker	180	67.0	Pria
25	Keith Flores	155	48.0	Pria
26	Carl Bryant	201	122.0	Pria
27	Sharon Cooper	162	62	Wanita
28	Denise Bryant	148	49	Wanita
29	Brian King	171	50.0	Pria
30	Rebecca Jenkins	196	86	Wanita
31	Alice Sanders	163	46	Wanita
32	Ashley Mitchell	159	57	Wanita
33	Anthony Anderson	163	53.0	Pria
34	Sarah Edwards	150	39	Wanita
35	Betty Griffin	170	56	Wanita
36	Peter Harris	191	55.0	Pria
37	Willie Lee	175	67.0	Pria
38	Kimberly Martin	169	78	Wanita
39	Julie Jones	167	59	Wanita
40	Albert Lopez	170	78.0	Pria
41	Benjamin Edwards	178	79.0	Pria
42	Donna Howard 168	168	71	Wanita
43	Marie Lewis	170	37	Wanita
44	Paula Kelly	167	58	Wanita
45	Irene Taylor	152	43	Wanita
46	Steve Clark	191	81.0	Pria
47	Barbara Hall	155	48	Wanita
48	Gerald Turner	176	61.0	Pria

No	Nama	Tinggi (cm)	Berat (Kg)	Jenis Kelamin
49	Evelyn Peterson	141	41	Wanita
50	Kelly Flores	166	59	Wanita
51	Craig Collins	168	46.0	Pria
52	Jessie Bell	165	65	Wanita
53	Douglas Griffin	169	67.0	Pria
54	Laura Rivera	158	43	Wanita
55	Joshua Davis	173	61.0	Pria
56	Raymond Kelly	180	74.0	Pria
57	Eric Price	212	59.0	
58	Marilyn Walker	152	62	Wanita
59	Shawn Nelson	189	67.0	Pria
60	Jessie Washington	159	56	Wanita
61	Diana Miller	163	58	Wanita
62	Jessie Davis	174	45	Wanita
63	Justin Richardson	174	69.0	Pria
64	Arthur Rivera	167	47.0	Pria
65	Sandra Johnson	131	37	Wanita
66	Ann Perry	154	74	Wanita
67	Helen Diaz	159	59	Wanita
68	Nicole Bailey	159	58	Wanita
69	Lisa Powell	177	83	Wanita
70	Adam Ross	193	96.0	Pria
71	Julia Brooks	180	83	Wanita
72	Fred Taylor	164	54.0	Pria
73	Louise Simmons	164	64	Wanita
74	George Evans	171	52.0	Pria
75	Anna Cox	163	41	Wanita
76	Jessie Johnson	165	30.0	Pria
77	Brenda Evans	161	61	Wanita
78	William Powell	198	75.0	Pria
79	Deborah Hughes	183	70	Wanita
80	Andrew Roberts	185	71.0	Pria
81	Richard Wilson	175	58.0	Pria
82	Harry Scott	195	89.0	Pria
83	Angela Price	170	66	Wanita
84	Sara Hill	167	61	Wanita
85	Maria Jackson	166	65	Wanita
86	Debra Gray	180	88	Wanita
87	Chris Perez	164	55.0	

No	Nama	Tinggi (cm)	Berat (Kg)	Jenis Kelamin
88	Lillian Henderson	161	53	Wanita
89	Edward Martin	187	76.0	Pria
90	Theresa Collins	170	63	Wanita
91	Donald Howard	192	101.0	Pria
92	Thomas Ramirez	175	56.0	Pria
93	Johnny Stewart	190	100.0	Pria
94	Matthew Rodriguez	164	63.0	Pria

No	Nama	Tinggi (cm)	Berat (Kg)	Jenis Kelamin
95	Phyllis Clark	172	61	Wanita
96	Jessie Thomas	168	69	Wanita
97	Emily Gonzalez	156	51	Wanita
98	Doris Nelson	167	40	Wanita
99	Louis Bennett	161	18.0	Pria
100	Janice Moore	167	56	Wanita

Hasil dari eksekusi terhadap data training adalah seperti berikut ini :

Wanita :

['Stephanie', 'Cynthia', 'Katherine', 'Elizabeth', 'Carol', 'Christina', 'Beverly', 'Sharon', 'Denise', 'Rebecca', 'Alice', 'Ashley', 'Sarah', 'Betty', 'Kimberly', 'Julie', 'Donna', 'Marie', 'Paula', 'Irene', 'Barbara', 'Evelyn', 'Kelly', 'Jessie', 'Laura', 'Marilyn', 'Jessie', 'Diana', 'Jessie', 'Sandra', 'Ann', 'Helen', 'Nicole', 'Lisa', 'Julia', 'Louise', 'Anna', 'Brenda', 'Deborah', 'Angela', 'Sara', 'Maria', 'Debra', 'Lillian', 'Theresa', 'Phyllis', 'Jessie', 'Emily', 'Doris', 'Janice']
 [149 174 183 138 145 161 179 162 148 196 163 159 150 170 169 167 168 170
 167 152 155 151 166 165 158 152 159 163 174 131 154 159 159 177 180 164
 163 161 183 170 167 166 180 161 170 172 168 156 167 167]

Pria :

['Randy', 'Jessie', 'David', 'Stephen', 'Jerry', 'Billy', 'Earl', 'Todd', 'Martin', 'Kenneth', 'Frank', 'Wayne', 'Henry', 'Alan', 'Paul', 'Jack', 'Bobby', 'Keith', 'Carl', 'Brian', 'Anthony', 'Peter', 'Willie', 'Albert', 'Benjamin', 'Steve', 'Gerald', 'Craig', 'Douglas', 'Joshua', 'Raymond', 'Eric', 'Shawn', 'Justin', 'Arthur', 'Adam', 'Fred', 'George', 'Jessie', 'William', 'Andrew', 'Richard', 'Harry', 'Chris', 'Edward', 'Donald', 'Thomas', 'Johnny', 'Matthew', 'Louis']
 [184 175 187 192 204 180 184 174 177 200 193 189 188 187 187 190 180 155
 201 171 163 191 175 170 178 191 176 168 169 173 180 212 189 174 167 193
 164 171 165 198 185 175 195 164 187 192 175 190 164 161]

3.2 Feature Class

Pada tahap pengujian ini, dibutuhkan label-label *Tinggi* dan *Nama_awal*. Label-label ini akan digabungkan dengan pengklasifikasian data tinggi badan pada interval [130..134, 135..139, 140..144, dan seterusnya]. Kemudian pengklasifikasian tinggi pria dan tinggi wanita dengan metode Naïve Bayes. hasil eksekusi terhadap klasifikasi data tersebut dapat dilihat dibawah ini.

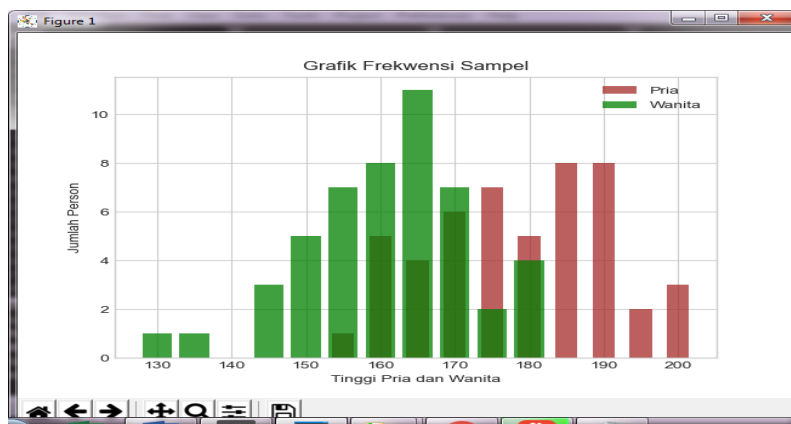
Pria {155: 1, 160: 5, 165: 4, 170: 6, 175: 7, 180: 5, 185: 8, 190: 8, 195: 2, 200: 3}

Wanita {130: 1, 135: 1, 140: 0, 145: 3, 150: 5, 155: 7, 160: 8, 165: 11, 170: 7, 175: 2, 180: 4, 185: 0}

Dengan memperhatikan data diatas, maka untuk pria yang memiliki tinggi tubuh 185 cm dan 190 cm akan muncul sebanyak 8 kali. Sedangkan untuk wanita adalah dengan yang bertubuh tinggi 165 cm dengan jumlah kemunculan sebanyak 11 kali.

3.3 Diagram Frekuensi Data

Adapun hasil eksekusi penghitungan frekuensi/jumlah data ntuk maasing-masing nilai atribut dan kelasnya sebagai berikut dapat dilihat seperti bentuk garafik dibawah ini:



Gambar 2. Grafik Data Frekuensi

Pada gambar di atas, terlihat bahwa jumlah/frekuensi kemunculan data tinngi wanita lebih sering muncul dibandingkan data tinggi pria.

3.4 Class Probabilitas

Perhitungan *Class Probabilities* adalah jumlah/frekuensi data untuk setiap nilai kelas dibagi dengan total jumlah/frekuensi data *training*. Seperti halnya seperti pada *Feature Class*, hasil yang sama juga didapatkan pada kelas probabilitas ini,

Pria {155: 1, 160: 5, 165: 4, 170: 6, 175: 7, 180: 5, 185: 8, 190: 8, 195: 2, 200: 3}

Wanita {130: 1, 135: 1, 140: 0, 145: 3, 150: 5, 155: 7, 160: 8, 165: 11, 170: 7, 175: 2, 180: 4, 185: 0}

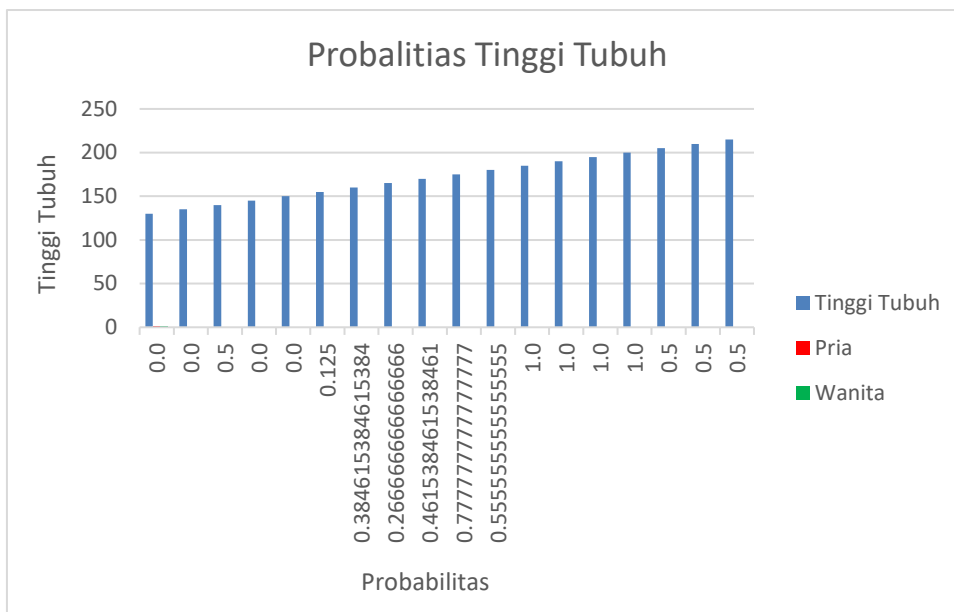
3.5 Pengujian terhadap kelas classifier

Merupakan tahap akhir dari pengujian terhadap tinggi tubuh dari data pada tabel 2 mulai dari 130 cm hingga 215 cm. Dan hasilnya adalah seperti berikut ini.

```

130 [(0.0, 'Pria'), (1.0, 'Wanita')]
135 [(0.0, 'Pria'), (1.0, 'Wanita')]
140 [(0.5, 'Pria'), (0.5, 'Wanita')]
145 [(0.0, 'Pria'), (1.0, 'Wanita')]
150 [(0.0, 'Pria'), (1.0, 'Wanita')]
155 [(0.125, 'Pria'), (0.875, 'Wanita')]
160 [(0.3846153846153847, 'Pria'), (0.6153846153846154, 'Wanita')]
165 [(0.26666666666666666, 'Pria'), (0.7333333333333333, 'Wanita')]
170 [(0.4615384615384616, 'Pria'), (0.5384615384615385, 'Wanita')]
175 [(0.7777777777777778, 'Pria'), (0.22222222222222224, 'Wanita')]
180 [(0.5555555555555556, 'Pria'), (0.4444444444444445, 'Wanita')]
185 [(1.0, 'Pria'), (0.0, 'Wanita')]
190 [(1.0, 'Pria'), (0.0, 'Wanita')]
195 [(1.0, 'Pria'), (0.0, 'Wanita')]
200 [(1.0, 'Pria'), (0.0, 'Wanita')]
205 [(0.5, 'Pria'), (0.5, 'Wanita')]
210 [(0.5, 'Pria'), (0.5, 'Wanita')]
215 [(0.5, 'Pria'), (0.5, 'Wanita')]
    
```

Adapun bentuk grafik dari data ouput di atas adalah seperti dibawah ini.



Gambar 3. Probabilitas tinggi tubuh

Gambar 3 menunjukkan bahwa tingkat probabilitas data tinggi tubuh wanita lebih besar daripada data tinggi tubuh pria. Jumlah total probabilitas tinggi tubuh pria mulai dari 130 cm hingga 215 adalah sebesar 8,6. Sedangkan probabilitas tinggi tubuh wanita = 9,6.

4. KESIMPULAN

Dari tahap penelitian sampai pada tahap pengujian implementasi dengan data inputan, penulis membuat beberapa kesimpulan:

1. Rancang bangun aplikasi berbasis desktop dengan menggunakan bahasa pemrograman Python versi 3.7.2.
2. Nilai-nilai atribut dan probabilitas yang ada pada metode Naive Bayes Classifier dapat dioptimalkan melalui proses prediksi suatu kondisi berdasarkan inputan pada klarifikasi data terkait.
3. Dari hasil pengujian terhadap aplikasi didapatkan bahwa dengan menggunakan algoritma *Naive Bayes Classifier*, maka frekuensi kemunculan data klasifikasi tinggi tubuh wanita lebih banyak daripada data klasifikasi tinggi tubuh pria.
4. Jumlah probabilitas tinggi tubuh wanita lebih besar dari jumlah probabilitas tinggi tubuh pria.

REFERENSI

[1] I. D. and A. S. Suasmini, *Penempatan Posisi Ketinggian Monitor Diturunkan Dapat Mengurangi Keluhan Subjektif Para Pemakai Kaca Bifokal*, vol. 2, no. 11. Denpasar: Institut Seni Indonesia, 2011.

[2] Anonim, "Pentingnya Pengukuran Antropometri dalam Bidang Kesehatan di Indonesia," 2019. [Online]. Available: <https://soloabadi.com/pentingnya-pengukuran-antropometri-dalam-bidang-kesehatan-di-indonesia/>. [Accessed: 21-Mar-2021].

[3] A. Irianton, *Memantau dan Menilai Status Gizi Anak*. Yogyakarta: Leutika Books, 2013.

[4] A. Istiany and Rusilanti, *Gizi Terapan*. Bandung: PT Remaja Rosdakarya, 2013.

[5] D. L. Olson and D. Delen, *Advanced Data Mining Techniques*. Verlag Berlin Heidelberg: Springer, 2008.

[6] D. Xhemali, C. J. Hinde, and R. G. Stone, "Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages," *Int. J. Comput. Sci.*, vol. 4, no. 1, pp. 16–23, 2009.

[7] Bustami, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," *TECHSI J. Penelit. Tek. Inform.*, vol. 3, no. 2, p. 127, 2013.

[8] A. Saleh, "Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citec J.*, vol. 2, no. 3, p. 207, 2015.

[9] I. Listiowarni, "Implementasi Naive Bayesian dengan Laplacian Smoothing untuk Peminatan dan Lintas Minat Siswa SMAN 5 Pamekasan," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 8, no. 2, p. 124, 2019.