

# METODE CLASSIFIER GAUSSIAN NAÏVE BAYES UNTUK KLASIFIKASI KONSUMSI MAKANAN CEPAT SAJI PADA TINGKAT RESIKO OBESITAS

Agus Nursikuwagus<sup>1</sup>, Suherman<sup>2</sup>

<sup>1,4</sup>Universitas Komputer Indonesia  
Jln. Dipatiukur No.112-114 Bandung – Indonesia

<sup>2</sup>Jurusan Sistem Informasi Fakultas Teknologi Informasi Universitas Serang Raya  
Jln. Raya Cilegon Serang – Drangong Kota Serang

<sup>1</sup>agusnursikuwagus@email.unikom.ac.id

<sup>2</sup>suherman.unsera@gmail.com

## Abstract

*The increasing prevalence of obesity, particularly among adolescents, has become a significant public health concern. Main contribution factors are the excessive consumption of fast food, which is often high in calories, fats, sodium, and sugar. This shift in dietary habits is largely leveraged by globalization, which promotes a more convenient yet unhealthy lifestyle. In response to this issue, this study focuses on the designation of fast-food consumption based on nutrition facts, specifically employing data from McDonald's menu items. The primary goal is to identify consumption patterns and predict whether a particular is at liability of dynamic obesity. This research utilizes the Gaussian Naïve Bayes (GNB) classification algorithm, implemented using the Python programming language. The dataset consists of 500 entries containing nutritional information such as calories, cholesterol, sodium, carbohydrates, sugars, protein, vitamins, calcium, fat, iron, fiber, potassium, minerals, and condition labels indicating obesity risk. The data was carved into training and testing sets, with 20% (100 data points) allocated for testing to assess the conduct of the model. The results of the experiment show that the GNB algorithm achieved an accuracy of 61%. For comparison, other machine learning models were also tested: Random Forest (RF) reach out the accuracy at 68%, succeeded by Multinomial Naïve Bayes at 65%, Support Vector Machine (SVM) and Decision Tree (DT) both at 62%. GNB showed limitations in predicting the TRUE class (indicating obesity risk), with a precision score of 60% and recall of 61%. These results suggest that the dataset distribution and the learning characteristics of GNB contributed to its relatively lower performance. Among the tested models, Random Forest demonstrated the best predictive capability, achieving a TRUE class precision of 76%, significantly outperforming GNB. This indicates that Random Forest is more reliable for classifying obesity risk based on fast food nutritional content. As a direction for future work, improvements can be made by tuning the parameters of the Random Forest model and exploring the application of deep learning techniques to further enhance predictive performance and generalizability.*

**Keywords:** Classifier, Gaussian Naive Bayes, Klasifikasi, Preprocessing, Modeling, Good Health and Well-being.

## I. PENDAHULUAN

Makanan *fast-food* merupakan makanan yang dikemas serta disajikan dengan cara cepat dan praktis serta siap disantap dengan waktu yang relatif singkat seperti menu yang disediakan di restoran cepat saji salah satunya yaitu McDonald's [1]. Akan tetapi makanan cepat saji ini umumnya mengandung lemak, garam, kalori, dan gula tinggi, namun rendah akan nutrisi vitamin, serat, folat dan kalsium [2]. Kehadiran makanan cepat saji pada beberapa restoran di Indonesia ini berpengaruh tinggi pada pola konsumsi makanan untuk sebagian masyarakat yang ada di Indonesia, khususnya pada remaja [3].

Oleh karena itu, kandungan nutrisi yang ada pada makanan cepat saji ini sangat berpengaruh bagi kesehatan manusia yang mengkonsumsinya. Selain itu, efek dari makanan cepat saji ini dapat berpengaruh kepada tingkat energi tubuh. Maka tidak dianjurkan untuk mengkonsumsi makanan cepat saji secara berlebihan [4]. Banyaknya masyarakat yang mengkonsumsi

makanan cepat saji dapat meningkatkan penimbunan kalori dalam tubuh yang nantinya akan menjadi penyebab peningkatan nilai IMT (Indeks Masa Tubuh) atau gizi berlebih [5].

Porsi makanan yang berenergi tinggi dan berukuran besar dapat memicu akumulasi kalori berlebih [6]. Kalori, lemak, serta natrium yang berlebihan dalam tubuh dapat menimbulkan dampak negatif bagi kesehatan dan memicu berbagai penyakit, misalnya obesitas atau kelebihan berat badan [7]. Kondisi obesitas terjadi ketika jaringan lemak dalam tubuh menumpuk secara signifikan [8]. Meskipun obesitas bukanlah penyakit langsung, kondisi ini menjadi faktor pemicu munculnya berbagai gangguan kesehatan seperti penyakit jantung, hipertensi, diabetes mellitus, serta penyakit degeneratif lainnya [9]. Penderita obesitas biasanya memiliki penimbunan lemak berlebihan pada jaringan otot, pankreas, maupun hati. Distribusi lemak yang tidak merata dapat berkembang menjadi bentuk seperti tumor [10].

Tidak sedikit masyarakat yang belum mengetahui status nutrisi yang terdapat pada *fast-food* ini. Dengan demikian, maka penulis akan melakukan pengelompokan atau pengklasifikasian kadar nutrisi pada makanan cepat saji ini dengan berdasarkan pada algoritma gaussian naïve bayes serta nantinya akan diimplementasikan pada database yang sudah ada kedalam jupyter notebook dan dibantu dengan bahasa pemrograman python [11]. Melihat penelitian-penelitian yang telah dilakukan sebelumnya, dan mempertimbangkan nutrisi makanan cepat saji yang terdapat pada menu mcdonalds dengan melakukan klasifikasi menggunakan metode-metode tersebut [10].

Kontribusi yang diberikan pada penelitian ini terkait penerapan pada prediksi obesitas antara lain :

- Mesin learning gaussian naïve bayes yang bekerja pada nilai data numerik dapat memberikan prediksi hasil klasifikasi dari dataset yang diberikan.
- Parameter dataset sebagai atribut yang diproses pada classifier gaussian memiliki nilai yang saling terkait antar atribut sehingga dapat menentukan prediksi dari obesitas.

Penyusunan bab dan subbab pada artikel ini dikelola dan disajikan agar mudah dipahami oleh pembaca berdasarkan urutan sebagai berikut : bagian satu pendahuluan berisikan mengenai problem dari state of the art dari perkembangan mesin learning. Bagian dua berisikan mengenai metode yang digunakan pada penelitian ini. Bagian tiga berisikan mengenai hasil dan pembahasan dari penelitian yang dikerjakan. Bagian empat merupakan kesimpulan dari rangkaian penelitian yang telah dikerjakan.

## II. METODOLOGI PENELITIAN

Secara etimologis, klasifikasi berarti pengelompokan atau penggolongan. Dalam pengertian umum, klasifikasi dipahami sebagai aktivitas menata pengetahuan secara menyeluruh ke dalam susunan yang sistematis [12]. Menurut Towa P. Hmakotrd dan J. N. B. Tairas (1995), klasifikasi merupakan proses pengelompokan yang terstruktur terhadap sejumlah objek, ide, buku, ataupun benda lain ke dalam kelas tertentu berdasarkan kesamaan karakteristik. Sementara itu, Suwarno (2007) menjelaskan bahwa klasifikasi adalah proses mengelompokkan objek atau barang sesuai tingkat kemiripannya [13].

### A. Fast-Food

Secara umum, makanan cepat saji digemari oleh remaja maupun orang dewasa karena rasanya yang sesuai dengan selera masyarakat serta harganya yang relatif ekonomis [14]. Istilah *fast-food* mencakup beragam hidangan yang proses persiapannya dilakukan dalam waktu singkat. Bertram (1975) dalam Hayati (2000) menjelaskan bahwa makanan cepat saji adalah hidangan yang disajikan secara praktis dan cepat [15]. Menurut Dr. Adhika Putra Rakhmatullah (2016), jenis *fast-food* yang paling populer di kalangan konsumen meliputi ayam goreng tepung, burger, dan kentang goreng [16].

### B. Obesitas

Obesitas adalah keadaan medis yang ditandai oleh berat badan berlebih yang disebabkan oleh penumpukan lemak

secara berlebihan di dalam tubuh [17]. Faktor pemicu utamanya adalah kebiasaan makan yang tidak sehat disertai asupan kalori yang melampaui kebutuhan, sehingga menyebabkan peningkatan massa tubuh [18]. Dampak obesitas tidak hanya dirasakan pada aspek fisik, tetapi juga dapat memengaruhi kesehatan mental (Hasdianah, 2014). Secara umum, kondisi ini dikategorikan sebagai gangguan patologis karena adanya penimbunan lemak yang berlebihan, yang berpotensi mengganggu fungsi kesehatan tubuh secara menyeluruh [19], [20].

Pengaruh genetik dari orang tua merupakan salah satu faktor yang memicu timbulnya obesitas pada anak [21]. Apabila kedua orang tua menderita obesitas, peluang anak mengalami kondisi yang sama dapat mencapai sekitar 80%. Namun, apabila hanya satu orang tua yang mengalami obesitas, peluangnya menurun menjadi sekitar 40%. Sementara itu, ketika kedua orang tua memiliki berat badan normal, kemungkinan anak mengalami kegemukan hanya sekitar 14% [22].

Dari sudut pandang biologis, makanan berperan sebagai sumber energi yang mengandung berbagai zat gizi penting, antara lain protein, karbohidrat, lemak, vitamin, serta mineral. Nutrien tersebut dimanfaatkan melalui proses metabolisme tubuh untuk mendukung pertumbuhan dan perbaikan sel, menjaga keseimbangan elektrolit, serta memperkuat sistem kekebalan tubuh [23].

### C. Gaussian Naïve Bayes (GNB)

Naïve Bayes merupakan metode klasifikasi yang sederhana namun populer, yang bekerja dengan menghitung probabilitas kelas berdasarkan penyebaran kata dalam sebuah dokumen [24]. Algoritma ini berlandaskan pada teorema Bayes dalam statistika, yang digunakan untuk memperkirakan probabilitas keanggotaan suatu data terhadap kelas tertentu [25]. Pendekatan Naïve Bayes memanfaatkan model probabilitas sampel untuk menghitung sekumpulan probabilitas yang berasal dari frekuensi dan kombinasi nilai pada dataset [26]. Bentuk umum dari teorema Bayes dapat dituliskan sebagai berikut:

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

Penjelasan masing-masing simbol adalah sebagai berikut:

- X: Data yang kelasnya belum ditentukan.
- H: Asumsi bahwa data termasuk pada kelas tertentu.
- P(H|X): Probabilitas posterior, yaitu kemungkinan hipotesis H benar dengan mempertimbangkan kondisi X.
- P(H): Probabilitas prior, yaitu kemungkinan awal terjadinya hipotesis H sebelum data X dianalisis.
- P(X|H): Peluang munculnya data X dengan anggapan hipotesis H benar.
- P(X): Peluang terjadinya data X secara umum.

Distribusi Gaussian Naïve Bayes merupakan salah satu metode sangat konvensional dan bermanfaat dalam perhitungan mengenai probabilitas dan statistik [27]. *Naïve Bayes* merupakan algoritma dengan menggunakan klasifikasi data berdasarkan perhitungan probabilitas kelompok dan menjumlahkan kombinasi nilai yang terdapat pada data yang telah dikumpulkan. Klasifikasi *Naïve Bayes* mempunyai hasil

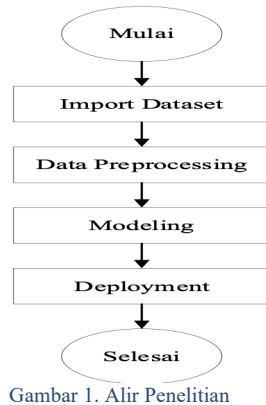
akurasi yang baik dan stabil. Namun, *Naïve Bayes* menghasilkan beberapa persyaratan yang menghasikan nilai akurasi yang rendah jika ada penambahan fitur dan parameter data. Persamaan pada formula *Gaussian Naïve Bayes*

$$\text{Classifier} : P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \bar{x})^2}{2\sigma^2}} \quad (2)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (3)$$

#### D. Alir Penelitian

Pada bagian ini menjelaskan alir penelitian yang digunakan ketika eksperimen. Berikut pada Gambar 1 diagram penelitian yang dilakukan.



Gambar 1. Alir Penelitian

#### E. Variabel Penelitian

Pada perancangan prediksi resiko penyakit obesitas dengan algoritma Gaussian Naïve Bayes Classifier ini menggunakan 14 variabel dengan parameter menu calories, cholesterol, sodium, carbohydrates, sugars, protein, vitamin, calcium, fat, iron, fiber, potassium, minerals, dan condition. Dengan menentukan 1 output yaitu “Pasien Resiko Terkena Penyakit Obesitas” atau “Pasien Tidak Beresiko Terkena Penyakit Obesitas”.

#### F. Python dan Streamlit

Python merupakan bahasa pemrograman dengan sifat multi-paradigma. Bahasa ini sepenuhnya mendukung pemrograman berorientasi objek serta pemrograman terstruktur, dan memiliki berbagai fitur yang memungkinkan penerapan pemrograman fungsional maupun pemrograman berorientasi aspek. Selain itu, melalui ekstensi, Python juga dapat digunakan untuk paradigma lain seperti desain berbasis kontrak dan pemrograman logika. Python mengadopsi sistem dynamic typing serta memanfaatkan kombinasi penghitungan referensi untuk pengelolaan memori [28].

Streamlit merupakan framework open-source yang dibangun menggunakan bahasa pemrograman Python, dirancang untuk mempermudah pembuatan aplikasi web, terutama di ranah data science dan machine learning, dengan menyediakan beragam fitur yang menunjang proses pengembangan model machine learning [29]. Platform ini bersifat gratis, dan pengguna tidak memerlukan keahlian

mendalam dalam pengembangan front-end untuk dapat menggunakannya. Streamlit dapat dijalankan pada editor seperti Anaconda dengan Python versi 3.8 atau lebih baru, namun tidak kompatibel secara langsung dengan Jupyter Notebook. Oleh karena itu, kode perlu dikonversi terlebih dahulu dan dijalankan melalui editor seperti PyCharm atau Visual Studio Code.

### III. HASIL DAN PEMBAHASAN

Pada subbab ini dituliskan hasil yang diperoleh menurut eksperimen yang dilakukan.

#### A. Dataset dan Data Preprocessing

Dalam penelitian ini menggunakan dataset dari kumpulan data fakta nutrisi makanan cepat saji yang dimiliki oleh McDonald’s yang kami temukan untuk diteliti. Terdapat 500 yang tersedia dalam dataset ini. Dataset ini memiliki 14 atribut yang terdiri dari Calories, Cholesterol, Sodium, Carbohydrates, Sugars, Protein, Vitamin, Calcium, Fat, Iron, Fiber, Potassium, Minerals, Condition.

Berdasarkan Gambar 2, memperlihatkan jumlah dataset fakta nutrisi dari sebuah makanan cepat saji yang sudah berhasil di running melalui jupyter notebook dengan jumlah rows 500 dan columns 13.

	calories	cholesterol	sodium	carbohydrates	sugars	protein	vitamin	
0	360	11.0	5.0	9.0	200	750	11	
1	250	0.0	3.0	0.0	25	770	30	
2	270	22.0	0.0	0.0	40	780	20	
3	400	22.0	16.0	0.0	205	800	30	
4	600	23.0	0.0	0.0	50	880	30	
...	...	...	...	...	...	...	...	
495	120	0.0	0.0	22.0	1	1	34	
496	100	0.0	0.0	43.0	1	1	40	
497	230	0.0	0.0	52.0	1	1	41	
498	230	0.0	0.0	89.0	1	1	46	
499	520	21.0	60.0	71.0	0	30	3	
...	...	...	...	...	...	...	...	
495	2	3	18	0	8	15		
496	3	3	18	4	0	15		
497	2	3	17	0	0	15		
498	3	4	17	4	0	15		
499	12	0	170	18	6	8		

Gambar 2. Dataset Resiko Obesitas Makanan Cepat Saji

Deskripsi atribut pada dataset Gambar 2, tertuang pada Tabel 1.

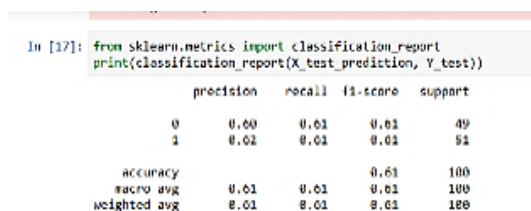
#### B. Modeling

Berdasarkan data uji yang digunakan untuk melakukan klasifikasi oleh tes dataset dengan metode Gaussian Naïve Bayes, pada Gambar 3 diperoleh nilai precision untuk kondisi bernilai 0 yang dimana “Pasien Tidak Beresiko Terkena Penyakit Obesitas” dengan perhitungan sebesar 0.60, nilai recall = 0.61, f1-score = 0.61 dan support = 40. Untuk kondisi bernilai 1 yang dimana “Pasien Beresiko Terkena Penyakit Obesitas” dengan hasil = 0.62, recall = 0.61, f1-score = 0.61 dan support = 51 instans. Serta total accuracy yang didapat dari hasil penelitian ini adalah sebesar 0.61.

TABLE I. ATTRIBUTE DATASET

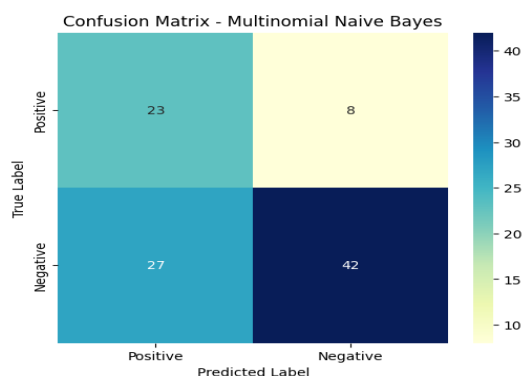
No	Attribute	Type	Descriptive
1	Calories	number	Kalori
2	Cholesterol	number	Kolesterol

3	Sodium	number	Natrium
4	Carbohydrates	number	Karbohidrat
5	Sugars	number	Gula
6	Protein	number	Protein
7	Vitamin	number	Vitamin
8	Calcium	number	Kalsium
9	Fat	number	Lemak
10	Iron	number	Zat Besi
11	Fiber	number	Serat
12	Potassium	number	Kalium
13	Minerals	number	Mineral
14	Condition	number	Kondisi



Gambar 3. Akurasi, Precision, Recall, dan F1-Measure GNB

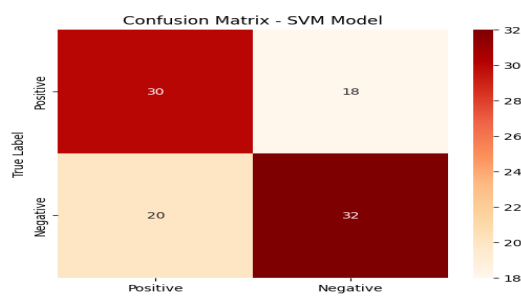
Model Multinomial Naive Bayes menghasilkan performa klasifikasi dengan True Positive (TP) = 23, dan True Negative (TN) = 42. Namun, model juga menghasilkan False Positive (FP) yang cukup tinggi, yakni 27, serta False Negative (FN) sebanyak 8. Dari hasil ini, akurasi model mencapai sekitar 65,3%, menunjukkan bahwa secara keseluruhan model benar dalam mengklasifikasikan sekitar dua pertiga data. Precision bernilai 46%, mengindikasikan bahwa hanya hampir setengah dari prediksi positif yang benar. Recall sebesar 74,2% mengindikasikan bahwa model memiliki kemampuan yang cukup baik dalam mendeteksi kasus positif. Namun, nilai F1-score sebesar 56,9% memperlihatkan bahwa keseimbangan antara precision dan recall masih belum maksimal. Tingginya tingkat error tipe I (false positive) menunjukkan bahwa cukup banyak kasus negatif yang salah terklasifikasi sebagai positif. Sebaliknya, tingkat error tipe II (false negative) relatif lebih rendah, yang berarti model lebih jarang gagal mendeteksi kasus positif. Hal ini menjadi signifikan terutama pada situasi di mana prioritas utama adalah mengidentifikasi kasus positif [30].



Gambar 4. Matrik konfusi Multinomial NB

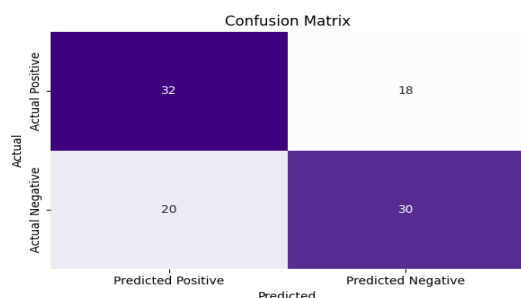
Support Vector Machine (SVM) yaitu algoritma machine learning yang dapat diterapkan baik untuk tugas regresi maupun klasifikasi. SVM bekerja dengan membangun sebuah

hyperplane yang memisahkan kelas-kelas data secara optimal. Dalam menangani data non-linier, SVM menggunakan teknik transformasi data ke dalam ruang berdimensi lebih tinggi agar pemisahan antar kelas menjadi lebih mudah dilakukan. Model Support Vector Machine (SVM) menghasilkan hasil klasifikasi dengan True Positive (TP) = 30 dan True Negative (TN) = 32. Sementara itu, False Positive (FP) = 20 dan False Negative (FN) = 18. Akurasi model tercatat sebesar 62%, yang menunjukkan kemampuan moderat dalam mengklasifikasikan data secara keseluruhan. Precision berada di angka 60%, artinya dari seluruh prediksi positif, 60% di antaranya benar-benar positif. Recall mencapai 62,5%, menandakan bahwa model berhasil mendeteksi lebih dari setengah kasus positif secara tepat. Nilai F1-score berada di kisaran 61%, mencerminkan keseimbangan yang cukup antara precision dan recall. Kesalahan tipe I (false positive) berjumlah 20, mengindikasikan bahwa sejumlah kasus negatif salah diklasifikasikan sebagai positif. Sementara itu, kesalahan tipe II (false negative) sebanyak 18 menunjukkan adanya kasus positif yang luput dari deteksi. Model ini cocok untuk konteks yang membutuhkan keseimbangan antara deteksi positif dan menghindari kesalahan klasifikasi.



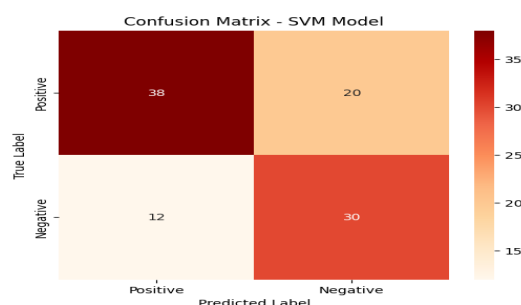
Gambar 5. Matrik konfusi support vector machine (SVM)

Decision Tree adalah salah satu algoritma machine learning yang dapat digunakan baik untuk tugas klasifikasi maupun regresi. Algoritma ini membentuk struktur pohon keputusan, di mana setiap node mewakili fitur tertentu dan setiap cabang menunjukkan nilai dari fitur tersebut. Tujuan utamanya adalah untuk membagi data ke dalam kelompok-kelompok homogen berdasarkan nilai fitur yang diberikan. Model Decision Tree menunjukkan performa yang relatif seimbang dengan jumlah True Positive (TP) = 31 dan True Negative (TN) = 31. Namun, model ini masih menghasilkan False Positive (FP) dan False Negative (FN) yang sama-sama berjumlah 19. Akurasi model mencapai sekitar 62%, yang menunjukkan bahwa hampir dua pertiga prediksi model sesuai dengan kelas sebenarnya. Precision sebesar 62% mencerminkan bahwa dari semua prediksi positif, sebagian besar benar. Sementara itu, recall sebesar 62% juga menandakan bahwa model mampu mendeteksi lebih dari separuh kasus positif. Nilai F1-score berkisar 62%, menandakan keseimbangan yang cukup antara presisi dan sensitivitas. Error tipe I (False Positive) menunjukkan 19 kasus negatif yang salah diklasifikasikan sebagai positif, sedangkan error tipe II (False Negative) juga sebanyak 19, yaitu kasus positif yang gagal dikenali. Keseimbangan antara kesalahan ini menunjukkan bahwa model Decision Tree bersifat netral, namun masih memiliki ruang untuk perbaikan [31].



Gambar 6. Matrik konfusi Decision Tree (DT)

Random Forest yaitu machine learning berbasis ensemble yang digunakan untuk pengkelasan dan regresi. Metode ini merekonstruksi banyak pohon keputusan secara paralel dan memberikan hasil prediksi berdasarkan agregasi (seperti voting atau rata-rata) dari semua pohon. Setiap pohon dibangun dari subset data dan fitur yang diambil secara acak. Random Forest sangat efektif dalam mengolah data berukuran besar dengan banyak fitur. Model Random Forest menghasilkan performa klasifikasi yang cukup solid dengan True Positive (TP) = 38 dan True Negative (TN) = 30. Meski begitu, masih terdapat False Positive (FP) = 12 dan False Negative (FN) = 20. Akurasi model mencapai 68%, menunjukkan tingkat prediksi yang baik secara keseluruhan. Precision sebesar 76% mencerminkan bahwa sebagian besar prediksi positif adalah benar, yang penting untuk menghindari alarm palsu. Recall = 65.5% dapat diinterpretasikan meskipun model bagus dalam mengenali kasus positif, namun masih ada beberapa yang terlewat. Nilai F1-score = 70.3% menunjukkan balancing antara kemampuan mengenali kasus positif dan akurasi prediksi tersebut. Error tipe I (FP) yang relatif kecil menunjukkan model jarang salah memberi label positif pada kasus negatif. Sementara itu, error tipe II (FN) yang lebih tinggi menunjukkan ada sejumlah kasus positif yang tidak terdeteksi. Model ini cocok digunakan dalam konteks di mana false alarm harus diminimalkan.



Gambar 7. Matrik konfusi Random Forest (RF)

Dari hasil evaluasi pada Tabel II terhadap lima algoritma klasifikasi, Random Forest menunjukkan performa terbaik secara umum dengan akurasi tertinggi sebesar 68%. Hal ini diperkuat dengan nilai presisi sebesar 76% yang mengindikasikan bahwa model ini sangat baik dalam meminimalkan kesalahan positif palsu (false positives). Nilai recall-nya pun relatif tinggi (66%), menunjukkan kemampuan yang baik dalam mendeteksi semua kelas positif. Sebaliknya, Multinomial Naïve Bayes memiliki karakteristik yang berbeda. Meskipun akurasinya cukup tinggi (65%), presisinya hanya sebesar 46%. Ini berarti banyak prediksi positif yang salah. Namun, nilai recall yang tinggi (74%) menandakan bahwa model ini sangat sensitif dalam mendeteksi kelas positif, namun

dengan konsekuensi menghasilkan banyak false positive. Hal ini dapat diterima tergantung konteks, terutama bila false negative lebih berbahaya [32], [33].

TABLE I. PERBANDINGAN KINERJA

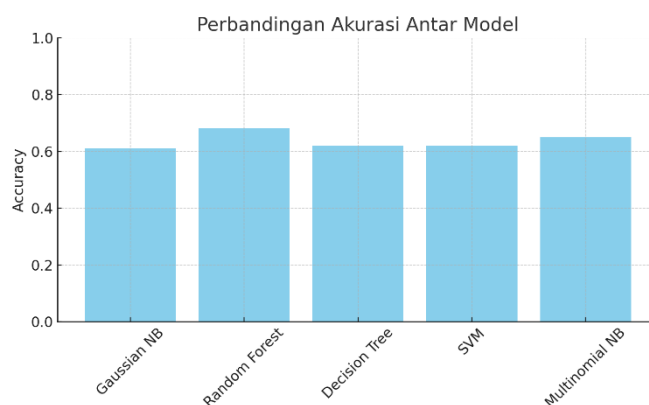
Model	Akurasi	Presisi	Recall	F1-Score
Gaussian NB	0.61	0.60	0.61	0.61
Random Forest	0.68	0.76	0.66	0.70
Decision Tree	0.62	0.62	0.62	0.62
SVM	0.62	0.60	0.63	0.61
Multinomial NB	0.65	0.46	0.74	0.57

Pada Gambar 8, Model Gaussian Naïve Bayes, SVM, dan Decision Tree memiliki performa yang mirip dengan akurasi sekitar 61–62% dan F1-score sekitar 0.61–0.62. Ketiganya memiliki keseimbangan antara presisi dan recall, meskipun tidak unggul secara signifikan dibanding Random Forest.

Dalam konteks aplikasi seperti deteksi penyakit, dimana false negative bisa sangat kritis, model dengan recall tinggi seperti Multinomial NB bisa lebih diprioritaskan. Namun, jika kestabilan performa secara keseluruhan lebih diutamakan, maka Random Forest menjadi pilihan yang paling andal.

### C. Deployment

Pengembangan selanjutnya yaitu deployment. Deployment ini adalah membuat aplikasi simulasi yang bisa diakses mandiri dengan bantuan Website. Gambar 9 merupakan website Prediksi Resiko Obesitas ini dibuat dengan bahasa pemrograman Python serta bantuan framework Streamlit dan metode algoritma Gaussian Naïve Bayes.



Gambar 8. Perbandingan Akurasi

### Prediksi Resiko Obesitas

Kalori

Kolesterol

Natrium

Karbohidrat

Gula

Protein

Vitamin

Kalsium

Lemak

Zat Besi

Serat

Kalium

Mineral

Prediksi Resiko Obesitas

Gambar 9. Website Prediksi Resiko Obesitas



Pada Gambar 10 ditunjukkan hasil prediksi yaitu “kondisi negatif”. Hasil prediksi kondisi pada website Prediksi Resiko Obesitas dengan jumlah Kalori sebesar 300, Kolesterol sebesar 260, Natrium sebesar 4, Karbohidrat sebesar 10, Gula sebesar 13, Protein sebesar 750, Vitamin sebesar 3, Kalsium sebesar 25, Lemak sebesar 5, Zat Besi sebesar 31, Serat sebesar 17, Kalium sebesar 15, dan Mineral 0 adalah “Pasien Tidak Beresiko Terkena Penyakit Obesitas”.

The screenshot shows a web application titled "Prediksi Resiko Obesitas". It features a form with input fields for various nutritional factors: Kalori (300), Kolesterol (260), Natrium (4), Karbohidrat (10), Gula (13), Protein (750), Vitamin (3), Kalsium (25), Lemak (5), Zat Besi (31), Serat (17), Kalium (15), and Mineral (0). Below the form is a button labeled "Prediksi Resiko Obesitas". The result displayed at the bottom is "Pasien Tidak Beresiko Terkena Penyakit Obesitas".

Gambar 10. Website Prediksi Resiko Obesitas Negatif

Pada Gambar 11 ditunjukkan hasil prediksi yaitu “kondisi positif”. Hasil prediksi kondisi pada website Prediksi Resiko Obesitas dengan jumlah Kalori sebesar 250, Kolesterol 0, Natrium sebesar 2, Karbohidrat sebesar 6, Gula 0, Protein 0, Vitamin sebesar 2, Kalsium sebesar 10, Lemak 0, Zat Besi sebesar 36, Serat sebesar 18, Kalium sebesar 20, dan Mineral 63 adalah “Pasien Resiko Terkena Penyakit Obesitas”.

The screenshot shows the same web application "Prediksi Resiko Obesitas" but with different input values: Kalori (250), Kolesterol (0), Natrium (2), Karbohidrat (6), Gula (0), Protein (0), Vitamin (2), Kalsium (10), Lemak (0), Zat Besi (36), Serat (18), Kalium (20), and Mineral (63). The result displayed at the bottom is "Pasien Resiko Terkena Penyakit Obesitas".

Gambar 11. Website Prediksi Resiko Obesitas Positif

Pada perolehan hasil akurasi masih dibawah 70%, hal ini masih perlu adanya tuning terhadap mesin learning yang digunakan. Nilai precisióń sebesar 0.61 menyatakan bahwa área pencarian prediksi dari dataset yang true prediction bisa mencapai 61% dari jumlah instans yang true prediction.

Deployment yang diimplementasikan dengan menggunakan Streamlit dan Python untuk memudahkan penggunaan mesin prediksi. Gambar 9, Gambar 10, dan Gambar 11 merupakan hasil implementasi dalam bentuk WEB. Sehingga pengguna dapat mudah untuk menggunakannya [34].

#### IV. KESIMPULAN

Hasil eksperimen diperoleh kesimpulan bahwa model yang dibuat menggunakan metode Gaussian Naïve Bayes mampu melakukan diagnose apakah seseorang tersebut resiko terkena penyakit obesitas atau tidak beresiko terkena penyakit obesitas. Dengan dilakukannya implementasi dari 500 dataset menggunakan parameter menu *calories*, *cholesterol*, *sodium*, *carbohydrates*, *sugars*, *protein*, *vitamin*, *calcium*, *fat*, *iron*, *fiber*, *potassium*, *minerals*, dan *condition* didapatkan tingkat akurasi data training sebesar 0.625 dari 400 data, dan tingkat akurasi data testing 0.61 dari 100 data. Adapun perolehan hasil akurasi prediksi dengan metode naive bayes yang menghasilkan nilai skor sebesar 0.61 precision, 0.61 recall, 0.61 f1-score, dan 100 support. Dibandingkan dengan mesin learning lainnya Random Forest memiliki Tingkat akurasi paling tinggi yaitu 0.68. Namun dari eksperimen yang dilakukan akurasi yang diperoleh masih di bawah 70% sehingga memerlukan perbaikan dari sisi mesin learning dan juga menambah dataset.

#### V. SARAN

Selanjutnya, untuk keberlanjutan penelitian maka dilakukan uji coba pada jenis algoritma klasifikasi data mining yang lainnya sehingga dapat dilakukan perbandingan metode dengan variasi model yang lebih sangkil dan mangkus untuk melakukan prediksi resiko terkena penyakit obesitas.

#### VI. TERIMAKASIH

Ucapan terima kasih disampaikan kepada seluruh mahasiswa mata kuliah Data Mining, khususnya Tia Eka Safitri, Vania Shinta Puspaningtyas, Fadilah Wulan Safitri, dan Angkasa Rachmat Susilo, yang telah berkontribusi dalam proses pengumpulan data serta pelaksanaan rangkaian implementasi dan deployment.

#### VII. DAFTAR PUSTAKA

- [1] A. E. Puspa, "Strategi Pengembangan Bisnis Produk Fried Chicken D'besto Di Pt Setyanda Duta Makmur," Skripsi, Fakultas Sains dan Teknologi, UIN Syarif Hidayatullah Jakarta, 2019.
- [2] U. Zein dan E. El Newi, Buku Ajar Ilmu Kesehatan (Memahami Gejala, Tanda Dan Mitos), Yogyakarta: Deepublish, 2019.
- [3] R. Nurdiansyah, "Budaya Pola Konsumsi Makanan Cepat Saji Dalam Kehidupan Remaja Jakarta (Studi Kasus: Franchise KFC)," Skripsi, Fakultas Ilmu Sosial dan Ilmu Politik, UIN Jakarta, 2019.
- [4] R. Muzakki, "Hubungan Konsumsi Makanan Cepat Saji Terhadap Peningkatan Risiko Obesitas Pada Anak Di Amerika Serikat."
- [5] L. Mulyana dan E. Farida, "Pola Pemberian Makan Yang Tepat Dalam Mengurangi Resiko Obesitas Pada Balita," Indonesian Journal of Public Health and Nutrition, vol. 2, no. 1, hlm. 36–42, 2022.
- [6] F. Z. Salsabila dan S. D. Puspawati, "Hubungan Konsumsi Fast Food Dengan Status Gizi Pada Remaja Di SMA Negeri 1 Kota Surakarta," Disertasi, Universitas Muhammadiyah Surakarta, 2022.
- [7] N. U. Tanjung, A. P. Amira, N. Muthmainah, dan S. Rahma, "Junk Food Dan Kaitannya Dengan Kejadian

- Gizi Lebih Pada Remaja," *Jurnal Ilmiah Kesehatan Masyarakat*, vol. 14, no. 3, hlm. 133–140, 2022.
- [8] T. Khoirunnisa dan R. Kurniasari, "Pengaruh Edukasi Melalui Media Pada Kejadian Overweight Dan Obesitas: Literature Review," *Prepotif: Jurnal Kesehatan Masyarakat*, vol. 6, no. 2, hlm. 1212–1217, 2022.
- [9] M. Pakpahan, dkk., *Pengantar Kesehatan Masyarakat*, Yayasan Kita Menulis, 2022.
- [10] W. Kusmawati, L. Lufthansa, R. S. Sari, dan S. M. Windriyani, *Buku Ajar Ilmu Gizi Olahraga*, Uwais Inspirasi Indonesia, 2019.
- [11] A. Senika, R. Rasiban, dan D. Iskandar, "Implementasi Metode Naïve Bayes Dalam Penilaian Kinerja Sales Marketing Pada Pt. Pachira Distrinusa," *Jurnal Media Informatika Budidarma*, vol. 6, no. 1, hlm. 701–709, 2022.
- [12] M. Muhajriah, "Analisis Penerapan Sistem Klasifikasi Dewey Decimal Classification (DDC) Di Dinas Perpustakaan Dan Kearsipan Kabupaten Enrekang," *Disertasi*, Universitas Islam Negeri Alauddin Makassar, 2020.
- [13] M. Rumra, Y. Pasoreh, dan R. Lesnusa, "Pengaruh Penentuan Nomor Klasifikasi Terhadap Temu Kembali Informasi Bagi Pemustaka Di Perpustakaan Universitas Muhammadiyah Maluku Utara," *Acta Diurna Komunikasi*, vol. 2, no. 4, 2020.
- [14] P. A. Mulyani, N. W. A. Sudiartini, dan N. L. P. Sariyani, "Perilaku Masyarakat Kota Denpasar Dalam Mengonsumsi Makanan Cepat Saji (Fast Food)," *Juima: Jurnal Ilmu Manajemen*, vol. 10, no. 2, 2020.
- [15] H. N. Putri, H. Nisyak, dan G. L. Elake, "Food Globalization: Dampak Kudapan Fast Food Terhadap Eksistensi Kuliner Khas Kota Palembang," *Disertasi*, Universitas Sriwijaya, 2021.
- [16] E. R. Eriska, A. K. Sari, dan L. N. Hakim, "Pengaruh Sistem Pelayanan, Kualitas Produk Terhadap Peningkatan Penjualan Di Restoran Cepat Saji Burger King," *Kalianda Halok Gagas*, vol. 5, no. 1, hlm. 14–22, 2022.
- [17] D. P. Delima dan R. T. Prasetyo, "Sistem Pakar Diagnosa Komplikasi Obesitas Pada Remaja Menggunakan Metode Certainty Factor," *Eprosiding Sistem Informasi (Potensi)*, vol. 2, no. 1, hlm. 54–60, 2021.
- [18] A. Darmawan, "Pengaruh Aktivitas Fisik Dan Pola Makan Sebagai Intervensi Penurunan Berat Badan Pada Anak Dengan Obesitas," *Disertasi*, Universitas Dr. Soebandi, 2022.
- [19] S. R. Amanda, "Hubungan Aktivitas Fisik Terhadap Obesitas Pada Anak Usia Dini."
- [20] N. M. J. S. N. Arini, "Hubungan Aktivitas Fisik Dengan Kejadian Obesitas Pada Remaja," 2020.
- [21] W. M. Wahid dan M. B. Arimbi, "Pengaruh Latihan Aerobik Terhadap Penurunan Ketebalan Lemak Subkutan," *Riyadhoh: Jurnal Pendidikan Olahraga*, vol. 4, no. 2, hlm. 63–70, 2022.
- [22] S. Manggabarani, A. J. Hadi, dan S. Ishak, "Edukasi Aktivitas Fisik Dalam Pencegahan Obesitas Di Madrasah Tsanawiyah Aisyiyah Kota Binjai," *Jurnal Pengabdian Masyarakat Ilmu Kesehatan*, vol. 1, no. 1, hlm. 1–6, 2020.
- [23] E. E. Mawarni, "Perubahan Fisiologis Yang Berhubungan Dengan Aspek Gizi Dan Pengaruhnya Pada Status Gizi Lansia," *Gizi Dalam Daur Kehidupan*, vol. 4.2, hlm. 35, 2022.
- [24] E. Indrayuni, "Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," *Jurnal Khatulistiwa Informatika*, vol. 7, no. 1, 2019.
- [25] A. Wanto, dkk., *Data Mining: Algoritma Dan Implementasi*, Yayasan Kita Menulis, 2020.
- [26] R. Amelia, D. Darmansah, N. S. Pratiwi, dan M. E. Purbaya, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada Twitter," *Jurikom: Jurnal Riset Komputer*, vol. 9, no. 2, hlm. 338–343, 2022.
- [27] N. A'ayunnisa, Y. Salim, dan H. Azis, "Analisis Performa Metode Gaussian Naïve Bayes Untuk Klasifikasi Citra Tulisan Tangan Karakter Arab," *Indonesian Journal of Data and Science*, vol. 3, no. 3, hlm. 115–121, 2022.
- [28] S. A. Firdaus, "Prediksi Penyakit Diabetes Dengan Metode Hyperbolic Support Vector Machine," *Disertasi*, Universitas Komputer Indonesia, 2020.
- [29] M. F. R. Maula, "Automated Valuation Model Untuk Estimasi Nilai Pasar Rumah Berbasis Jaringan Saraf Tiruan Backpropagation," *Disertasi*, UIN Sunan Ampel Surabaya, 2022.
- [30] R. Prathivi dan S. Sharifzadeh, "Performance Evaluation of Naive Bayes Algorithm for Classification of Fertilizer Types," *Telematika*, vol. 15, no. 1, hlm. 57–69, Feb. 2022.
- [31] U. Modeling dan G. I. Webb, "Using Decision Trees for Agent Modeling: Improving Prediction Performance," *Mar.* 1998.
- [32] J. Liang, "Confusion Matrix: Machine Learning," *POGIL Activity Clearinghouse*, vol. 3, no. 4, 2022.
- [33] M. Heydarian, T. E. Doyle, dan R. Samavi, "MLCM: Multi-label Confusion Matrix," *IEEE Access*, vol. 10, hlm. 19083–19095, 2022.
- [34] R. Yadav dan H. Raheman, "Machine Learning-based Estimation of Agricultural Tyre Sinkage: A Streamlit Web Application," *J. Terramechanics*, vol. 119, p. 101055, 2025.