

Analisis Perbandingan KNN, SVM, Decision Tree dan Regresi Logistik Untuk Klasifikasi Obesitas Multi Kelas

Siti Andini Utiaharman^{1*}, A. Mulawati M. Pratama²

¹ Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Ichsan Gorontalo, Gorontalo, Indonesia

² Fakultas Teknologi dan Sains, Program Studi Sistem Informasi, Universitas Ichsan Gorontalo Utara, Gorontalo, Indonesia

Email: ^{1*}siti_andini@unisan.ac.id, ²mulapratama@gmail.com

Email Penulis Korespondensi: andiniutiaharman@gmail.com

Abstrak—Obesitas menjadi masalah kesehatan dunia yang mengkhawatirkan, dengan prevalensi yang terus meningkat. Identifikasi dini dan klasifikasi obesitas yang akurat menjadi sangat penting untuk menerapkan strategi pencegahan dan penanganan yang tepat. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa empat algoritma klasifikasi populer yaitu K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree dan Regresi Logistik dalam melakukan klasifikasi obesitas multi kelas berdasarkan Indeks Massa Tubuh (BMI) menurut WHO. Dengan menggunakan dataset yang mencerminkan keragaman populasi, penelitian ini mengevaluasi kemampuan masing-masing algoritma dalam mengklasifikasikan obesitas ke dalam beberapa katagori, seperti normal, overweight, obesitas tingkat 1, obesitas tingkat 2, dan obesitas tingkat 3. Jumlah data yang digunakan sebanyak 2.111 record dengan 17 atribut. Hasil penelitian menunjukkan bahwa algoritma Decision Tree mengungguli algoritma lainnya dengan mendapatkan hasil akurasi 99.3%, presisi 0.97-1.00, recall 0.98-1.00 dan f1-score 0.98-1.00. Disusul oleh KNN mendapatkan hasil akurasi 99.0%, presisi 0.98-1.00, recall 0.98-1.00 dan f1-score 0.98-1.00. Sementara itu algoritma Regresi Logistik mendapatkan akurasi 98%, presisi 0.95-1.00, recall 0.95-1.00 dan f1-score 0.95-1.00 dan SVM memiliki performa sedikit lebih rendah meskipun masih menunjukkan hasil yang baik secara keseluruhan dengan hasil akurasi 96.6%, presisi 0.90-0.99, recall 0.94-1.00 dan f1-score 0.93-0.99.

Kata Kunci: Analisis; Decision Tree; KNN; Klasifikasi Multi Kelas; Obesitas; Regresi Logistik; SVM

Abstract—Obesity has become a concerning global health issue, with continuously increasing prevalence. Early identification and accurate classification of obesity are crucial for implementing appropriate prevention and treatment strategies. This study aims to analyze and compare the performance of four popular classification algorithms: K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree, and Logistic Regression, in performing multi-class obesity classification based on Body Mass Index (BMI) according to World Health Organization (WHO) standards. Using a dataset reflecting population diversity, this research evaluates the ability of each algorithm to classify obesity into several categories, such as normal, overweight, obesity grade 1, obesity grade 2, and obesity grade 3. The study utilizes 2.111 records with 17 attributes. Results indicate that the Decision Tree Algorithm outperforms other algorithms, achieving an accuracy of 99.3%, precision of 0.97-1.00, recall of 0.98-1.00, and f1-score of 0.98-1.00. KNN follows with an accuracy of 99.0%, precision of 0.98-1.00, recall of 0.98-1.00 and f1-score of 0.98-1.00. meanwhile, the Logistic Regression algorithm achieves an accuracy of 98%, precision of 0.95-1.00, recall of 0.95-1.00, and f1-score of 0.95-1.00. SVM demonstrates slightly lower performance, although still showing overall good results with an accuracy of 96.6%, precision of 0.90-0.99, recall of 0.94-1.00, and f1-score of 0.93-0.99.

Keywords: Analysis; Decision Tree; KNN; Logistic Regression; Multiclass Classification; Obesity; SVM

1. PENDAHULUAN

Obesitas menjadi tantangan kesehatan masyarakat dunia. Berdasarkan Data terbaru *World Health Organization (WHO)* pada tahun 2022, sebanyak 2.5 miliar orang dewasa berusia 18 tahun keatas mengalami kelebihan berat badan dan 890 termasuk menderita obesitas [1]. Jika tingkat pertumbuhan ini bertahan, proporsi orang yang masuk dalam katagori obesitas akan meningkat pada tahun 2030 [2]. Di Indonesia sendiri berdasarkan hasil Survei Kesehatan Indonesia 2023, angka obesitas di Indonesia meningkat menjadi 23,4% [3]. Obesitas dianggap sebagai faktor utama munculnya berbagai macam penyakit seperti jantung koroner, kanker, stroke iskemik, kardiovaskular, diabetes melitus tipe 2 dan penyakit lainnya. [4] [5]. Beberapa penelitian menyebutkan obesitas dapat mempengaruhi keparahan di pasien Covid-19 [6] [7]. Obesitas juga berkontribusi dalam meningkatnya biaya perawatan kesehatan dan adanya penurunan kualitas hidup [8]. Oleh karenanya, identifikasi dini dan klasifikasi obesitas menjadi sangat penting untuk menerapkan strategi pencegahan dan penanganan yang tepat.

Obesitas adalah keadaan tubuh seseorang dikala melebihi berat normal yang disebabkan penimbunan jaringan lemak tubuh berlebih. Seseorang yang obesitas bisa dengan mudah di ukur dengan menggunakan Indeks Massa Tubuh (BMI) [9]. Seseorang yang memiliki $IMT > 25,0$ dapat dikatakan kelebihan berat badan dan termasuk dalam katagori obesitas [10]. Dalam upaya mencegah obesitas, telah dilakukan berberapa penelitian untuk mengembangkan metode klasifikasi yang akurat dan efisien. Salah satu pendekatan yang sekarang populer adalah dengan mengimplementasikan metode *machine learning*. *Machine learning* terbukti efektif dalam menganalisis data kompleks dan menemukan pola tersembunyi [2]. Beberapa penelitian terdahulu telah menerapkan algoritma *machine learning* seperti *K-Nearest Neighbor (KNN)*, *Support Vector Machine (SVM)*, *Neural Network*, *Naive Bayes (NB)* untuk klasifikasi obesitas.

Penelitian oleh Wulandari dkk, menganalisis tingkat obesitas menggunakan teknik data mining dengan 3 algoritma ANN, KNN dan NB. Algoritma KNN memiliki akurasi lebih tinggi 95,74% [11]. Selanjutnya Putri dkk, mengimplementasikan algoritma KNN, NB dan SVM untuk mengklasifikasikan data obesitas. Namun tidak dijelaskan bagaimana karakteristik obesitas diidentifikasi. Akurasi tertinggi dihasilkan *Decision Tree* 84,98% [12]. Fitriani dan Bahri, melakukan pendekatan biner untuk membedakan antara individu yang termasuk dalam 2 katagori. Metode jaringan

saraf tiruan dilakukan dengan membagi dataset menjadi *training* dan data *testing* 70:30. Kinerja model dievaluasi menggunakan metrik evaluasi standar menghasilkan seperti akurasi 0.9684, presisi 0.9669, *recall* 0.9915 dan *F1-score* 0.9791. Katagori target diolah menjadi 2 katagori saja [13]. Selanjutnya penelitian oleh Hidayatulloh dan Yusuf, Klasifikasi tipe berat badan tubuh menggunakan metode SVM. Tujuan mengenali kelebihan berat badan seseorang dengan lebih cepat menggunakan algoritma SVM menggunakan data 252 *record* dan mencari tingkat akurasi dari 2 kelas klasifikasi yaitu normal dan kelebihan berat badan. Akurasi yang dihasilkan sebesar 92,11% dengan nilai kurva ROC sebesar 0,990 [14]. Jumlah data yang digunakan belum mencerminkan keragaman populasi secara menyeluruh, sehingga diperlukan data yang lebih besar. Penelitian Sibi dan Widiarti, mengimplementasikan variasi nilai *k* pada KNN, uji validasi silang, dan modifikasi atribut. Hasil penelitian menunjukkan bahwa model klasifikasi obesitas optimal menggunakan KNN dengan $k = 1$ dan atribut berat badan serta usia memiliki pengaruh yang signifikan. Akurasi maksimal diperoleh 79,96% [15]. Terakhir Wie dan Siddik, memprediksi obesitas pada pria pada menggunakan pengukuran data antropologi dan NB. Data yang digunakan terdiri dari 252 observasi dengan 15 variabel, dengan hanya pada subjek laki-laki karena massa massa ototnya lebih tinggi dan kecenderungan penumpukan jaringan adiposa di area perut. Hasil uji klasifikasi dengan rasio pelatihan-pengujian 70:30, menunjukkan metrik kinerja yang menjanjikan untuk algoritma NB. Akurasi 88%, presisi 73% untuk obesitas dan 98% untuk non-obesitas, *recall* 96% untuk obesitas dan 85% untuk non-obesitas dan *F1-score* 0,83. Penelitian ini hanya membatasi subjek laki-laki sehingga membatasi kemampuan generalisasi temuan [16].

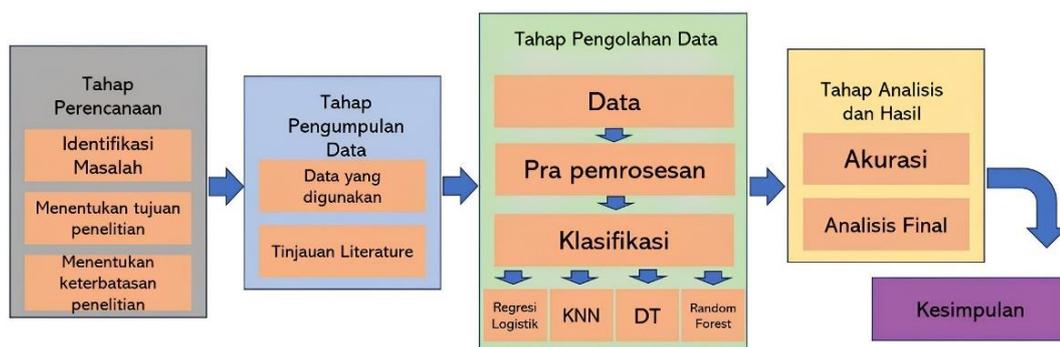
Meskipun penelitian-penelitian tersebut memberikan kontribusi berharga dalam pengembangan metode klasifikasi obesitas, masih terdapat beberapa kesenjangan yang perlu dicapai. Salah satunya adalah kurangnya perbandingan komprehensif antara berbagai algoritma klasifikasi dalam klasifikasi obesitas multi kelas berdasarkan Indeks Massa Tubuh (BMI) seperti yang ditunjukkan pada tabel 1 [17]. Selain itu, terdapat penelitian terdahulu berfokus pada klasifikasi biner (obesitas dan tidak obesitas) dan menggunakan dataset dengan ukuran terbatas yang mungkin kurang mencerminkan keragaman populasi secara menyeluruh.

Penelitian ini bertujuan untuk menganalisis dan membandingkan performa empat algoritma klasifikasi populer, yaitu KNN, SVM, *Decision Tree* dan Regresi Logistik dalam melakukan klasifikasi obesitas multi kelas. Dengan membandingkan keempat algoritma tersebut, penelitian ini akan mengevaluasi kemampuan masing-masing dalam mengklasifikasikan obesitas kedalam beberapa katagori berdasarkan matrix evaluasi seperti akurasi, presisi, *recall* dan *f1-score*. Hasil penelitian diharapkan dapat mengidentifikasi algoritma yang paling akurat dan efisien untuk klasifikasi obesitas multi kelas, serta memberikan rekomendasi terbaik untuk diimplementasikan dalam pengaplikasian manajemen berat badan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan desain eksperimental dan mengikuti metodologi spesifik yang ditunjukkan pada gambar 1.



Gambar 1. Gambaran tahapan penelitian

Tahap perencanaan diawali dengan melakukan identifikasi masalah, menentukan tujuan penelitian, studi kelayakan dan menentukan fokus dan batasan penelitian. Tahap kedua adalah mengumpulkan data yang akan digunakan. Data pada penelitian ini menggunakan data publik di website. Selanjutnya adalah mencari literatur terkait baik dari jurnal, buku maupun artikel relevan yang tersedia di website *kaggle.com*. Tahap ketiga adalah tahapan pengolahan data dengan melakukan prapemrosesan data sehingga menghasilkan data yang siap untuk diolah atau diklasifikasikan dengan empat model klasifikasi yaitu KNN, SVM, *Decision Tree* dan Regresi Logistik. Tahap keempat adalah tahapan memulai analisis dan melihat hasil dari analisis dengan cara memataui hasil akurasi yang dihasilkan oleh masing-masing model, termasuk hasil presisi, *recall* dan *F1-score*. Selanjutnya adalah menentukan analisis final. Terakhir adalah menarik kesimpulan dari hasil analisis. Klasifikasi BMI berdasarkan skema WHO ditunjukkan pada tabel 1.

2.2 Obesitas

Banyak orang menderita masalah kelebihan berat badan, yaitu masalah obesitas, tanpa mengetahui cara memeriksa obesitas, dan Indeks Massa Tubuh (BMI). Obesitas memiliki beberapa tingkatan, yaitu tingkat 1, 2 dan 3. Tingkatan tersebut telah ditentukan oleh BMI, yang pada gilirannya bergantung pada berat badan dan tinggi badan saja. Namun obesitas juga bergantung pada usia dan jenis kelamin. Misalnya, jika umur dua orang adalah 88 dan 22, sedangkan berat badan kedua umurnya sama, maka BMI kedua orang tersebut sama. Namun tingkat obesitasnya berbeda-beda untuk kedua orang tersebut. Sepengetahuan kami belum ada rumusan matematis yang demikian menjelaskan atau menghitung kesenjangan tingkat obesitas ini. Oleh karena itu dapat memanfaatkan teknik *machine learning* untuk mencapai hasil yang akurat berdasarkan nilai obesitas dalam berbagai situasi [18].

Tabel 1. Obesitas dalam skema BMI

BMI (kg/m x m)	Klasifikasi berat	Kelas Obesitas	Resiko Penyakit
<18.5	Kurus	-	-
18.5-24.9	Normal	-	-
25.0-29.9	Gemuk	-	Meningkat
30.0-34.9	Gendut	Obesitas tingkat 2	Tinggi
35.0-39.9	Obesitas	Obesitas tingkat 3	Sangat Tinggi
≥ 40	Obesita yang Ekstrim	Obesitas tingkat 4	Sangat Ekstrim

2.3 K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor* (KNN) merupakan teknik katagori data yang menemukan mana pengelompokkan suatu titik data yang memungkinkan suatu titik data akan bergabung kelompok itu. Salah satu contoh algoritma *lazy-learning* yaitu KNN [19], yang hanya memperkirakan fungsinya secara lokal dan tidak menyelesaikan semua perhitungan sampai tahap klasifikasi. Menemukan pengelompokkan k item pada data latih yang paling mirip dengan objek pada uji atau data baru adalah bagaimana KNN dijalankan keluar. Langkah-langkah yang diperlukan untuk menghitung KNN adalah sebagai berikut [11]:

- Menentukan parameter k
- Menghitung jarak antara data latih dan data uji
- Perhitungan jarak yang paling umum digunakan dalam perhitungan pada algoritma KNN adalah menggunakan perhitungan jarak Euclidean. Rumusnya ditunjukkan pada persamaan (1).

$$euc = \sqrt{(\sum_{i=1}^n (p_i - q_i)^2)}$$
 (1)

Rumus jarak *Euclidean* ini untuk mengukur kedekatan antar titik data. Rumus ini menghitung akar kuadrat dari jumlah selisih kuadrat antara fitur-fitur yang bersesuaian dari dua titik data. Jika diterapkan pada klasifikasi obesitas multi kelas, KNN akan menghitung jarak antara titik data yang ingin diklasifikasikan dengan semua titik data dalam dataset pelatihan. Selanjutnya algoritma ini memilih k tetangga terdekat dan mengklasifikasikan titik data baru berdasarkan mayoritas kelas dari tetangga-tetangga tersebut. Pendekatan ini memungkinkan dari kasus-kasus yang paling mirip dalam dataset.

2.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu algoritma *machine learning* yang efektif untuk mengakomodasi banyak variabel dan multi kelas [20]. SVM merupakan teknik pembelajaran terawasi yang mungkin mengidentifikasi pola dalam data dengan menganalisisnya. Analisis regresi dan klasifikasi memanfaatkannya. Itu konsep dasar dibalik pendekatan ini yaitu memilih ruang pemisah yang optimal dari kumpulan data multikelas. Katagorikal ini dilakukan dengan kelas lainnya [21] [22]. SVM ditunjukkan pada persamaan (2).

$$f(x_d) = \sum_{i=1}^{ns} a_i Y_i X_i X_d + b$$
 (2)

Rumus diatas menggunakan fungsi keputusan untuk melakukan klasifikasi. a_i adalah bobot *Lagrange*, Y_i adalah label kelas, X_i adalah *support vector*, X_d adalah titik data yang diklasifikasikan, dan b adalah bias. SVM bekerja dengan mencari *hyperplane* optimal yang memisahkan kelas-kelas obesitas dengan margin maksimal. *Support vector*, yang merupakan titik-titik data terdekat dengan *hyperplane*, memainkan peran kunci dalam menentukan batas keputusan. Untuk klasifikasi obesitas multi kelas, SVM dapat menggunakan pendekatan *one-vs-one*, dimana *multiple hyperplane* digunakan untuk memisahkan setiap kelas obesitas dari kelas lainnya.

2.5 Decision Tree (DT)

Decision Tree merupakan algoritma pohon hierarki yang dibuat dengan mempartisi data menjadi beberapa set sesuai untuk memasukan variabel sebagai pendekatan *data mining* untuk klasifikasi item. *Decision Tree* dibagi lagi menjadi beberapa tipe antara lain ID3, CART, C4.5 dan lainnya. Algoritma ID3 yang dibuat oleh Qiuilan, diperluas oleh algoritma C4.5. ID3 merupakan metode yang digunakan untuk data katagorikal, sedangkan C4.5 digunakan untuk kedua katagorikal

dan data numerik. Sementara C4.5 menggunakan rasio gain sebagai kriteria pemilihan atributnya, teknik ID3 digunakan perolehan informasi sebagai kriterianya [12]. Tahapan algoritma *Decision Tree* adalah sebagai berikut :

- a. Siapkan dataset pelatihan
- b. Temukan akar pohon keputusan
- c. Tentukan fitur dengan menghitung nilai gain yang akan dijadikan akar pohon keputusan. Itu nilai perolehan terbesar diantara kualitas yang tersedia digunakan untuk menghitung perolehan. Nilai keuntungannya mungkin dihitung menggunakan persamaan (3) berikut :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^N \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (3)$$

- d. Prosedurnya untuk setiap cabang yang terbentuk, ulangi langkah kedua. Disisi lain untuk menghitung nilai entropi, gunakan persamaan yang sesuai yang dapat menggunakan persamaan (4) sebagai berikut :

$$entropy(S) = \sum_{i=1}^N -\pi \times \log 2\pi \quad (4)$$

Decision Tree menggunakan persamaan (3) dan (4) untuk membangun struktur pohon keputusan. *Information gain* mengukur seberapa baik suatu atribut dalam memisahkan data ke dalam kelas-kelas obesitas yang berbeda. *Entropy* mengukur ketidakpastian dalam dataset, dengan nilai yang lebih rendah menunjukkan dataset yang lebih homogen. Pada klasifikasi obesitas, *Decision Tree* akan memilih atribut-atribut seperti (BMI, usia ataupun faktor gaya hidup) yang paling efektif dalam membedakan tingkat obesitas. Algoritma ini secara rekursif membagi dataset berdasarkan atribut-atribut terpilih, membentuk struktur pohon yang dapat digunakan untuk mengklasifikasikan kasus obesitas baru.

2.6 Regresi Logistik

Regresi logistik merupakan salah satu metode *machine learning* yang handal digunakan untuk klasifikasi data dengan bertipe katagori [20] Regresi logistik adalah jenis algoritma *machine learning* klasifikasi statistik. Dasar dari algoritma ini diturunkan dari fungsi *sigmoid*. Regresi logistik sangat berguna dalam masalah klasifikasi biner. Kelemahan dari regresi logistik adalah rentan terhadap *underfitting* dataset yang kelasnya tidak seimbang dapat sehingga menghasilkan akurasi yang rendah [23]. Regresi logistik dengan persamaan (5) :

$$P(Y = 1|X) = 1/(1 + e^{(-z)}) \quad (5)$$

$P(Y = 1|X)$ merupakan probabilitas bahwa variabel dependen (Y) adalah 1 (katagori positif), e merupakan bilangan *Euler* (sekitar 2.71828), z merupakan fungsi linier yang mengkombinasikan variabel independen (X) dengan koefisien regresi (β). Fungsi z dengan persamaan (6) berikut ini :

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_k X_k \quad (6)$$

β_0 merupakan *intercept*. $\beta_1, \beta_2, \dots, \beta_k$ adalah koefisien regresi yang mengukur seberapa besar pengaruh masing-masing variabel independen pada probabilitas katagori positif. X_1, X_2, \dots, X_k merupakan nilai-nilai variabel independen. Angka 0 menunjukkan suatu peristiwa yang tidak mungkin terjadi dan angka 1 menunjukkan kemungkinan maksimum bahwa peristiwa ini akan terjadi. Persamaan logistik menggunakan fungsi logaritma untuk menghitung garis regresi [24].

Untuk menghitung probabilitas kasus dalam kelas obesitas. Fungsi *sigmoid* ini memetakan kombinasi linear dari fitur-fitur input (seperti berat badan, tinggi badan dan faktor-faktor lain) ke dalam probabilitas antara 0 dan 1. Koefisien β menunjukkan kontribusi relatif dari setiap fitur terhadap prediksi tingkat obesitas. Untuk klasifikasi obesitas multi-kelas, regresi logistik dapat diperluas menggunakan teknik *one-vs-one*, dimana probabilitas untuk setiap kelas obesitas dihitung, dan kelas dengan probabilitas tertinggi dipilih sebagai hasil klasifikasi.

2.7 Akurasi Klasifikasi Multikelas

Akurasi pada klasifikasi adalah ukuran dari ketepatan klasifikasi dimana menunjukkan performansi teknik klasifikasi berbasis holistik. Semakin tinggi akurasi klasifikasi maka dianggap performansi teknik klasifikasi juga semakin baik [25]. Umumnya klasifikasi hanya melibatkan dua kelas, meskipun dapat lebih dari dua.

Dalam klasifikasi multi kelas, akurasi metrik yang mengukur seberapa baik model dapat memprediksi kelas yang benar dari total data yang ada. Dalam multi-kelas, variabel respon Y dan prediksi \hat{Y} dapat dianggap sebagai variabel acak diskrit yang mengambil nilai dalam $\{1, \dots, K\}$, dimana setiap angka mewakili kelas yang berbeda. Model klasifikasi memberikan probabilitas bahwa suatu unit data tertentu berada di salah satu kelas yang mungkin. Kemudian, aturan klasifikasi digunakan untuk menetapkan satu kelas pada setiap unit data. Cara mengukur kinerja klasifikasi menggunakan *confusion matrix*. *Confusion matrix* digunakan untuk mengevaluasi kinerja model klasifikasi multi-kelas. Contoh *confusion matrix* multi-kelas ditunjukkan pada tabel 2 [26].

Tabel 2. Contoh confusion matrix multi-kelas

	Predicted Classification				Total	
	Classes	a	b	c		d
Actual Classification	a	6	0	1	2	9
	b	3	9	1	1	14
	c	1	0	10	2	13

	d	1	2	1	12	16
Total	11	11	13	17	52	

Ukuran evaluasi kinerja model didasarkan *confusion matrix* dimana terdiri dari akurasi, *precision* dan *F1-score*. Akurasi dalam *confusion matrix* adalah presentase kebenaran dari prediksi pada data uji. *Precision* adalah ukuran proporsi prediksi Tru Positive. Akurasi, *precision*, *recall* dan *F1-score* diperoleh sebagai persamaan yang ditunjukkan pada (7), (8), (9), dan (10).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{7}$$

Mengacu pada tabel 1 maka :

$$Accuracy = \frac{6+9+10+12}{52}$$

$$Precision = \frac{TP}{FP+TP} \tag{8}$$

Mengacu pada tabel 1 maka :

$$Precision = \frac{20}{30} = 0.66$$

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

Mengacu pada tabel 1 maka :

$$Recall = \frac{20}{25} = 0.80$$

$$F1-Score = \frac{2xPrecision \times Recall}{Precision+Recall} \tag{10}$$

Mengacu pada tabel 1 maka :

$$F1-Score = \frac{2x0.66 \times 0.80}{0.66+0.80} = 0.72$$

3. HASIL DAN PEMBAHASAN

3.1 Identifikasi Data

Perkiraan tingkat obesitas pada rentang usia antara 14 sampai 61 tahun yang terdiri dari 2.111 *record* dengan 17 atribut yaitu : *Age*, *Gender*, *height*, *weight*, *CALC* (konsumsi alkohol), *FAVC* (konsumsi makanan berlemak), *FCVC* (konsumsi sayur/buah), *NCP* (makan diluar), *SCC* (merokok), *smoke*, *CH20* (konsumsi air putih), *family history with everweight*, *FAF* (aktifitas fisik), *TUE* (penggunaan teknologi), *CAEC* (konsumsi kalori), *MTRANS* (moda transportasi). Potongan dari data ditunjukkan pada tabel 3.

Target multi kelas pada klasifikasi obesitas adalah *Nobeyesdad* yang terdiri atas *Insufficient weight* (berat kurang dibawah berat normal), *Normal weight* (berat badan ideal), *Overweight level 1* (kelebihan berat badan tingkat 1), *Obesity type 1* (obesitas tipe ringan), *Obesity type II* (obesitas tipe sedang) dan *obesity type III* (obesitas berat) yang ditunjukkan pada gambar 2.

Tabel 3. Potongan informasi data yang digunakan

No	Age	Gender	Height	Weight	CALC	FAVC	FCVC	NCP	SCC	SMOKE	CH20	family_histoy _with_overweig ht	FAF	TUE	CAEC	MTRANS	NObesyedad
0	21	Female	1.62	64	no	no	2	3	no	no	2	yes	0	1	Sometime s	Public_Transportati on	Normal_Weight
1	21	Female	1.52	56	Sometime s	no	3	3	yes	yes	3	yes	3	0	Sometime s	Public_Transportati on	Normal_Weight
2	23	Male	1.8	77	Frequently	no	2	3	no	no	2	yes	2	1	Sometime s	Public_Transportati on	Normal_Weight
3	27	Male	1.8	87	Frequently	no	3	3	no	no	2	no	2	0	Sometime s	Walking	Overweight_Level I
4	22	Male	1.78	89.8	Sometime s	no	2	1	no	no	2	no	0	0	Sometime s	Public_Transportati on	Overweight_Level I
...
2106	20.97664 2	Female	1.71073	131.4085	Sometime s	yes	3	3	no	no	1.72813	yes	1.676269	0.906247	Sometime s	Public_Transportati on	Obesity_Type_III
2107	21.98294 2	Female	1.74858	133.7429	Sometime s	yes	3	3	no	no	2.00513	yes	1.34139	0.59927	Sometime s	Public_Transportati on	Obesity_Type_III
2108	22.52403 6	Female	1.75226	133.6894	Sometime s	yes	3	3	no	no	2.05419	yes	1.414209	0.646288	Sometime s	Public_Transportati on	Obesity_Type_III
2109	24.35409 4	Female	1.7595	133.3466	Sometime s	yes	3	3	no	no	2.852339	yes	1.139107	0.586035	Sometime s	Public_Transportati on	Obesity_Type_III
2110	23.66477	Female	1.73886	133.4726	Sometime s	yes	3	3	no	no	2.863513	yes	1.026452	0.714137	Sometime s	Public_Transportati on	Obesity_Type_III

```
target_unique
array(['Normal_Weight', 'Overweight_Level_I', 'Overweight_Level_II',
      'Obesity_Type_I', 'Insufficient_Weight', 'Obesity_Type_II',
      'Obesity_Type_III'], dtype=object)
```

Gambar 2. Kelas target

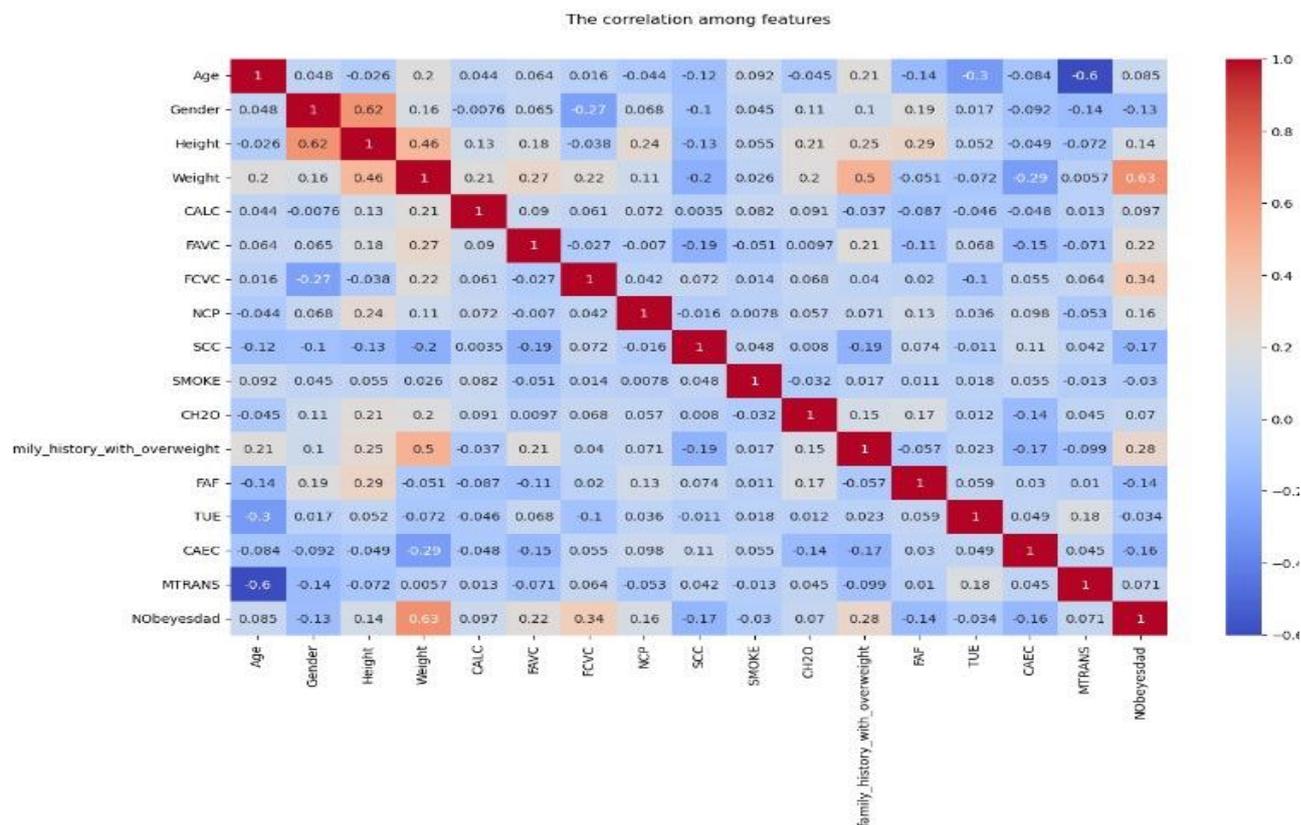
3.2 Pra Pemrosesan Data

Sebelum membangun model, dilakukan pra pemrosesan data seperti melihat hasil matriks korelasi antara berbagai fitur yang menunjukkan kekuatan dan arah hubungan antara setiap fitur. Hasilnya ditunjukkan pada gambar 3.

Berdasarkan hasil matriks korelasi didapatkan hubungan antar fitur sebagai berikut: Matriks korelasi menunjukkan hubungan antara 16 fitur yang terkait dengan obesitas. Nilai dalam matriks berkisar antara -1 dan 1, dengan nilai positif menunjukkan hubungan positif, nilai negatif menunjukkan hubungan negatif, dan 0 menunjukkan tidak ada hubungan.

Interpretasi Penting:

- Usia: Berhubungan positif lemah dengan semua fitur kecuali NObesyedad (indikator obesitas). Implikasinya: usia merupakan faktor risiko untuk semua jenis penyakit.
- Jenis Kelamin: Berhubungan negatif lemah dengan CALC, FAVC, FCVC (indikator risiko). Implikasinya: wanita mungkin kurang rentan terhadap faktor risiko ini dibandingkan pria.
- Tinggi Badan: Berhubungan positif lemah dengan CALC, FAVC, FCVC. Implikasinya: orang yang lebih tinggi mungkin lebih rentan terhadap faktor risiko ini.
- Berat Badan: Berhubungan positif lemah dengan CALC, FAVC, FCVC, NCP. Implikasinya: orang yang kelebihan berat badan mungkin lebih rentan terhadap faktor risiko ini.



Gambar 3. Matriks korelasi

3.3 Implementasi Algoritma

Pada implementasi ini kami menggunakan algoritma dioptimalkan dalam library seperti scikit-learn di Python Untuk validasi kinerja model menggunakan parameter accuracy, dan confusion matrix sebagai berikut :

- K-Nearest Neighbor (KNN)

Pada confusion matrix menunjukkan bahwa model KNN secara akurat mengklasifikasikan sebagian besar data point di semua kelas. Namun terdapat sedikit kesalahan klasifikasi antara kelas 2 dan 3, serta antara kelas 4 dan 5. Hasil confusion matrix model KNN ditunjukkan pada gambar 4.

```
confusion_matrix(y, best_knn_model.predict(x_norm) )
array([[281,  2,  1,  2,  1,  0,  0],
       [ 3, 284,  1,  2,  0,  0,  0],
       [ 0,  2, 286,  2,  0,  0,  0],
       [ 2,  0,  0, 349,  0,  0,  0],
       [ 2,  0,  2,  0, 268,  0,  0],
       [ 0,  0,  0,  0,  0, 297,  0],
       [ 0,  0,  0,  0,  0,  0, 324]], dtype=int64)
```

Gambar 4. Confusion matrix KNN

Model KNN yang dilatih dengan parameter yang diberikan menunjukkan kinerja yang sangat baik dalam mengklasifikasikan data menjadi 7 kelas. Model ini mencapai akurasi tinggi sebesar 99% dan metrik evaluasi yang baik untuk semua kelas dengan presisi tinggi untuk semua kelas 0.98-1.00, *recall* tinggi untuk semua kelas 0.98-1.00 dan *f1-score* tinggi untuk semua kelas 0.98-1.00. Hasil klasifikasi KNN ditunjukkan pada gambar 5.

```
print(classification_report(y, best_knn_model.predict(x_norm)))
```

	precision	recall	f1-score	support
2.0	0.98	0.98	0.98	287
3.0	0.99	0.98	0.98	290
4.0	0.99	0.99	0.99	290
5.0	0.98	0.99	0.99	351
6.0	1.00	0.99	0.99	272
7.0	1.00	1.00	1.00	297
8.0	1.00	1.00	1.00	324
accuracy			0.99	2111
macro avg	0.99	0.99	0.99	2111
weighted avg	0.99	0.99	0.99	2111

Gambar 5. Hasil klasifikasi KNN

b. Support Vector Machine (SVM)

Confusion matrix menunjukkan bahwa model SVM mengalami lebih banyak kesalahan klasifikasi dibandingkan model sebelumnya. Kesalahan klasifikasi terbanyak terjadi pada kelas 2 sebanyak 13 kesalahan dan kelas 3 sebanyak 10 kesalahan. Hasil *confusion matrix Support Vector Machine* ditunjukkan pada gambar 6.

```
confusion_matrix(y, best_svm_model.predict(x_norm) )
array([[255, 13,  0,  0, 19,  0,  0],
       [ 2, 278, 10,  0,  0,  0,  0],
       [ 0, 17, 272,  1,  0,  0,  0],
       [ 0,  0,  4, 344,  0,  3,  0],
       [ 1,  0,  0,  0, 271,  0,  0],
       [ 0,  0,  0,  2,  0, 295,  0],
       [ 0,  0,  0,  0,  0,  0, 324]], dtype=int64)
```

Gambar 6. Confusion Matrix SVM

Model SVM yang dilatih dengan parameter yang diberikan menghasilkan akurasi sebesar 96.6% lebih rendah dibandingkan model sebelumnya pada data *training*. Model ini memiliki performa sedikit lebih rendah pada metrik evaluasi dengan presisi tinggi untuk sebagian besar kelas 0.90-0.99, *recall* tinggi untuk sebagian besar kelas 0.94-1.00 dan *f1-score* tinggi untuk sebagian besar 0.93-0.99. Hasil klasifikasi SVM ditunjukkan pada gambar 7.

```
print(classification_report(y, best_svm_model.predict(x_norm)))
```

	precision	recall	f1-score	support
2.0	0.99	0.89	0.94	287
3.0	0.90	0.96	0.93	290
4.0	0.95	0.94	0.94	290
5.0	0.99	0.98	0.99	351
6.0	0.93	1.00	0.96	272
7.0	0.99	0.99	0.99	297
8.0	1.00	1.00	1.00	324
accuracy			0.97	2111
macro avg	0.97	0.96	0.96	2111
weighted avg	0.97	0.97	0.97	2111

Gambar 7. Hasil klasifikasi SVM

c. Decision Tree

Pada *confusion matrix* menunjukkan bahwa model *Decision Tree* mengklasifikasikan sebagian besar data point dengan akurat disemua kelas. Namun terdapat sedikit kesalahan klasifikasi antara kelas 2 dan 3, kelas 4 dan 5 serta kelas 6 dan 7. Hasil *confusion matrix* model *Decision Tree* ditunjukkan pada gambar 8.

```
confusion_matrix(y, best_DT_model.predict(x_norm) )
array([[283,  3,  0,  0,  1,  0,  0],
       [ 1, 289,  0,  0,  0,  0,  0],
       [ 0,  5, 284,  1,  0,  0,  0],
       [ 0,  0,  0, 351,  0,  0,  0],
       [ 1,  0,  0,  0, 271,  0,  0],
       [ 0,  0,  0,  2,  0, 295,  0],
       [ 0,  0,  0,  0,  0,  0, 324]], dtype=int64)
```

Gambar 8. Confusion matrix Decision Tree

Model *Decision Tree* yang dilatih dengan parameter yang diberikan menghasilkan akurasi tertinggi mencapai 99.3%. Model ini juga mencapai metrik evaluasi presisi tinggi untuk semua kelas 0.97-1.00, *recall* tinggi untuk semua kelas 0.98-1.00 dan *f1-score* tinggi untuk semua kelas 0.98-1.00. Hasil *confusion matrix* model *Decision Tree* ditunjukkan pada gambar 9.

```
print(classification_report(y, best_DT_model.predict(x_norm)))
```

	precision	recall	f1-score	support
2.0	0.99	0.99	0.99	287
3.0	0.97	1.00	0.98	290
4.0	1.00	0.98	0.99	290
5.0	0.99	1.00	1.00	351
6.0	1.00	1.00	1.00	272
7.0	1.00	0.99	1.00	297
8.0	1.00	1.00	1.00	324
accuracy			0.99	2111
macro avg	0.99	0.99	0.99	2111
weighted avg	0.99	0.99	0.99	2111

Gambar 9. Hasil klasifikasi Decision Tree

d. Regresi Logistik

Untuk hasil *confusion matrix* menunjukkan bahwa model secara akurat mengklasifikasikan sebagian besar data point disemua kelas. Namun terdapat sedikit kesalahan klasifikasi antara kelas 2 dan 3, serta antara kelas 4 dan 5. Hasil *confusion matrix* model Regresi Logistik ditunjukkan pada gambar 10.

```
confusion_matrix(y, best_logreg_model.predict(x_norm) )
array([[277,  2,  0,  0,  8,  0,  0],
       [ 4, 276, 10,  0,  0,  0,  0],
       [ 0, 13, 276,  1,  0,  0,  0],
       [ 0,  0,  1, 350,  0,  0,  0],
       [ 1,  0,  0,  0, 271,  0,  0],
       [ 0,  0,  0,  0,  0, 297,  0],
       [ 0,  0,  0,  0,  0,  0, 324]], dtype=int64)
```

Gambar 10. Confusion matrix Regresi Logistik

Untuk hasil klasifikasi dihasilkan akurasi ditunjukkan pada gambar 11.

```
print(classification_report(y, best_logreg_model.predict(x_norm)))
```

	precision	recall	f1-score	support
2.0	0.98	0.97	0.97	287
3.0	0.95	0.95	0.95	290
4.0	0.96	0.95	0.96	290
5.0	1.00	1.00	1.00	351
6.0	0.97	1.00	0.98	272
7.0	1.00	1.00	1.00	297
8.0	1.00	1.00	1.00	324
accuracy			0.98	2111
macro avg	0.98	0.98	0.98	2111
weighted avg	0.98	0.98	0.98	2111

Gambar 11. Hasil klasifikasi Regresi Logistik

Berdasarkan gambar 11, model *Regresi Logistik* yang dilatih dengan parameter yang diberikan menunjukkan kinerja yang sangat baik dalam mengklasifikasikan data menjadi 7 kelas. Model ini mencapai akurasi tinggi sebesar 98% dan

metrik evaluasi yang baik. Dimana presisi tinggi untuk semua kelas yaitu 0.95-1.00, *recall* tinggi untuk semua kelas yaitu 0.95-1.00. begitu juga *f1-score* tinggi untuk semua kelas 0.95-1.00

Pada hasil analisis yang diberikan, berikut adalah analisa akhir dari setiap model klasifikasi:

a. *K-Nearest Neighbor (KNN)*

1. Model KNN menunjukkan kinerja yang sangat baik dalam mengklasifikasikan data ke dalam 7 kelas dengan akurasi tinggi 99%.
2. Memiliki presisi, *recall*, dan *f1-score* yang tinggi untuk semua kelas (0.98-1.00).
3. Terdapat sedikit kesalahan klasifikasi antara kelas 2 dan 3, serta kelas 4 dan 5.

b. *Support Vector Machine (SVM)*

1. Model SVM menghasilkan akurasi lebih rendah (96.6%) dibandingkan model lainnya.
2. Memiliki presisi, *recall*, dan *f1-score* yang sedikit lebih rendah, terutama pada kelas 2 dan 3.
3. Mengalami lebih banyak kesalahan klasifikasi dibandingkan model lainnya.

c. *Decision Tree*

1. Model *Decision Tree* menghasilkan akurasi tertinggi (99.3%) di antara semua model yang diuji.
2. Memiliki presisi, *recall*, dan *f1-score* yang sangat tinggi untuk semua kelas (0.97-1.00).
3. Terdapat sedikit kesalahan klasifikasi antara kelas 2 dan 3, kelas 4 dan 5, serta kelas 6 dan 7.

d. Regresi Logistik

1. Model Regresi Logistik menunjukkan kinerja yang sangat baik dengan akurasi 98%.
2. Memiliki presisi, *recall*, dan *f1-score* yang tinggi untuk semua kelas (0.95-1.00).
3. Terdapat sedikit kesalahan klasifikasi antara kelas 2 dan 3, serta kelas 4 dan 5.

Secara keseluruhan, model *Decision Tree* dan KNN menunjukkan kinerja terbaik dengan akurasi tertinggi dan metrik evaluasi yang sangat baik untuk semua kelas. Sementara model SVM memiliki kinerja yang sedikit lebih rendah dibandingkan model lainnya.

3.4 Pembahasan

Berdasarkan hasil implementasi keempat model maka kami membandingkan dengan penelitian sebelumnya, hasil penelitian kami menunjukkan performa yang lebih baik dalam klasifikasi obesitas multi kelas dengan Akurasi tertinggi pada algoritma *Decision Tree* dengan nilai 99.03%, presisi 97%, *recall* 98% dan *f1-score* 98%. Dapat ditunjukkan pada tabel 3.

Tabel 4. Perbandingan dengan penelitian sebelumnya

No.	Penelitian	Metode	Multikelas	Performa Terbaik			
				Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
1.	[11]	ANN, KNN , NB	√	95.74%	95.71%	95.56%	-
2.	[12]	KNN, MB, SVM, DT	-	84.98%	95.41%	97.01%	-
3.	[13]	NN	-	96.84%	90.06%	97.32%	93.58%
4.	[14]	SVM	-	92.11%	-	-	-
5.	[15]	KNN	√	79.96%	-	-	-
6.	[16]	NB	√	88%	73%	96%	0.830
7.	kami	KNN, SVM, DT , RL	√	99.03%	97%	98%	98%

4. KESIMPULAN

Penelitian bertujuan untuk membandingkan performa empat algoritma *machine learning* populer, yaitu KNN, SVM, *Decision Tree* dan Regresi Logistik dalam melakukan klasifikasi obesitas multi kelas. Obesitas telah menjadi masalah kesehatan masyarakat global yang semakin mengkhawatirkan, dan klasifikasi yang akurat menjadi penting untuk menerapkan strategi pencegahan dan penanganan yang tepat. Dengan menggunakan dataset yang mencerminkan keragaman populasi, penelitian ini mengevaluasi kemampuan masing-masing algoritma dalam mengklasifikasikan obesitas ke dalam beberapa katagori berdasarkan BMI menurut WHO. Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* mengungguli algoritma lainnya dengan mendapatkan hasil akurasi 99.3%, presisi 0.97-1.00, *recall* 0.98-1.00 dan *f1-score* 0.98-1.00. Disusul oleh KNN mendapatkan hasil akurasi 99.0%, presisi 0.98-1.00, *recall* 0.98-1.00 dan *f1-score* 0.98-1.00. Sementara itu algoritma Regresi Logistik mendapatkan akurasi 98%, presisi 0.95-1.00, *recall* 0.95-1.00 dan *f1-score* 0.95-1.00 dan SVM memiliki performa sedikit lebih rendah meskipun masih menunjukkan hasil yang baik secara keseluruhan dengan hasil akurasi 96.6%, presisi 0.90-0.99, *recall* 0.94-1.00 dan *f1-score* 0.93-0.99. Dibandingkan dengan penelitian sebelumnya, hasil penelitian ini menunjukkan performa yang lebih baik dalam klasifikasi obesitas multi kelas, baik dalam hal akurasi, presisi, *recall* maupun *f1-score*. Namun, perlu diakui bahwa penelitian ini memiliki keterbatasan, diantaranya tidak mempertimbangkan faktor-faktor lain yang mungkin mempengaruhi klasifikasi seperti kesehatan individu dan genetik. Untuk penelitian selanjutnya disarankan dapat mengeksplorasi pengaruh faktor tersebut dan menggabungkannya dengan algoritma *machine learning* terbaik yang ditemukan dalam penelitian ini. Selain itu, penerapan algoritma terbaik dalam pengaplikasian manajemen berat badan untuk dikembangkan lebih lanjut.

REFERENCES

- [1] World Health Organization, "Obesity and overweight." <https://www.who.int>. Accessed: May 10, 2024. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>
- [2] B. Singh and H. Tawfik, "Machine learning approach for the early prediction of the risk of overweight and obesity in young people," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, 2020, pp. 523–535. doi: 10.1007/978-3-030-50423-6_39.
- [3] Direktorat Jenderal Pencegahan dan Pengendalian Penyakit (P2P) Kementerian Kesehatan RI, "Ancaman di Balik Sensasi Manis," <https://p2p.kemkes.go.id/>. Accessed: May 10, 2024. [Online]. Available: <https://p2p.kemkes.go.id/ancaman-di-balik-sensasi-manis/#:~:text=Menurut%20hasil%20Survei%20Kesehatan%20Indonesia,dapat%20menyebabkan%20munculnya%20resistensi%20insulin>.
- [4] B. R. K. Salim, D. M. Wihandani, and N. N. A. Dewi, "Obesitas sebagai faktor risiko terjadinya peningkatan kadar trigliserida dalam darah: tinjauan pustaka," *Intisari Sains Medis*, vol. 12, no. 2, pp. 519–523, Jul. 2021, doi: 10.15562/ism.v12i2.1031.
- [5] K. Jindal, N. Baliyan, and P. S. Rana, "Obesity prediction using ensemble machine learning approaches," in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2018, pp. 355–362. doi: 10.1007/978-981-10-8636-6_37.
- [6] A. Adimara, K. Prahastanti, and M. P. Airlangga, "Obesitas Mempengaruhi Tingkat Keparahan Pasien COVID-19," 2021.
- [7] A. Y. Soeroto et al., "Effect of increased BMI and obesity on the outcome of COVID-19 adult patients: A systematic review and meta-analysis," *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, vol. 14, no. 6. Elsevier Ltd, pp. 1897–1904, Nov. 01, 2020. doi: 10.1016/j.dsx.2020.09.029.
- [8] F. B. C. Pimenta, E. Bertrand, D. C. Mograbi, H. Shinohara, and J. Landeira-Fernandez, "The relationship between obesity and quality of life in Brazilian adults," *Front Psychol*, vol. 6, Jul. 2015, doi: 10.3389/fpsyg.2015.00966.
- [9] M. J. Gerl et al., "Machine learning of human plasma lipidomes for obesity estimation in a large population cohort," *PLoS Biol*, vol. 17, no. 10, 2019, doi: 10.1371/journal.pbio.3000443.
- [10] G. Alifa Annurullah et al., "Faktor Resiko Obesitas Pada Pekerja Kantoran : A Review," *Jurnal Kesehatan Tambusai*, vol. 2, no. 2, 2021.
- [11] A. Wulandari, A. Mulya, T. Dermawan, R. R. Haiban, A. Tatamara, and H. D. Khalifah, "Application of Artificial Neural Network, K-Nearest Neighbor and Naive Bayes Algorithms for Classification of Obesity Risk Cardiovascular Disease," *IJATIS: Indonesian Journal of Applied Technology and Innovation Science*, vol. 1, no. 1, pp. 9–15, 2024, doi: 10.57152/IJATIS.v1i1.1095.
- [12] A. I. Putri et al., "Implementation of K-Nearest Neighbors, Naïve Bayes Classifier, Support Vector Machine and Decision Tree Algorithms for Obesity Risk Prediction," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 2, no. 1, pp. 26–33, Apr. 2024, doi: 10.57152/predatecs.v2i1.1110.
- [13] D. Nur Fitriani et al., "Prediction of Obesity Levels Using Neural Network: Binary Classification Approach," *PARAMETER Jurnal Matematika Statistika dan Terapannya*, vol. 3, no. 1, pp. 85–88, 2024, [Online]. Available: <https://ojs3.unpatti.ac.id/index.php/parameter>
- [14] T. Hidayatulloh and L. Yusuf, "Klasifikasi Tipe Berat Tubuh Menggunakan Metode Support Vector Machine," *INTI Nusa Mandiri*, vol. 18, no. 1, pp. 71–77, Aug. 2023, doi: 10.33480/inti.v18i1.4254.
- [15] S. Y. Sibi and A. R. Widiarti, "Klasifikasi Tingkat Obesitas Mempergunakan Algoritma KNN," in *SEMINAR NASIONAL CORISINDO*, 2022, pp. 370–375.
- [16] J. V. Wie and M. Siddik, "Penerapan Metode Naive Bayes Dalam Mengklasifikasi Obesitas Pada Pria," *JOISIE Journal Of Information System And Informatics Engineering*, vol. 6, no. Desember, pp. 69–77, 2022, [Online]. Available: <https://www.kaggle.com/>,
- [17] M. Safaei, E. A. Sundararajan, M. Driss, W. Boulila, and A. Shapi'i, "A systematic Literature Review on Obesity: Understanding The Causes & Consequences of Obesity and Reviewing Various Machine Learning Approaches used to Predict Obesity," *Comput Biol Med*, vol. 136, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104754.
- [18] K. Jindal, N. Baliyan, and P. S. Rana, "Obesity prediction using ensemble machine learning approaches," in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2018, pp. 355–362. doi: 10.1007/978-981-10-8636-6_37.
- [19] M. Ramadhani, D. Darlis, and H. Murti, "Klasifikasi Ikan Menggunakan Oriented Fast And Rotated Brief (ORB) dan K-Nearest Neighbor (KNN)," *Jurnal Ilmiah Teknologi Informasi*, vol. 16, no. 2, pp. 115–124, 2018.
- [20] I. M. K. Karo, M. F. M. Fudzee, S. Kasim, and A. A. Ramli, "Sentiment Analysis in Karonese Tweet using Machine Learning," *Indonesian Journal of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 219–231, Mar. 2022, doi: 10.52549/ijeei.v10i1.3565.
- [21] R. Syahputra, G. J. Yanris, and D. Irmayani, "SVM and Naïve Bayes Algorithm Comparison for User Sentiment Analysis on Twitter," *Sinkron*, vol. 7, no. 2, pp. 671–678, May 2022, doi: 10.33395/sinkron.v7i2.11430.
- [22] et al Suryani, "View of Sentiment Analysis of Towards Electric Cars using Naive Bayes Classifier and Support Vector Machine Algorithm," *Journal of Engineering, Data Technology and Computer Science*, vol. 1, no. 1, pp. 1–9, Jul. 2023.
- [23] Hendriyana, "Tampilan Analisis perbandingan Algoritma Support Vector Machine, Naive Bayes dan Regresi Logistik untuk Memprediksi Donor Darah," *Jurnal Teknologi Terpadu*, vol. 8, no. 2, pp. 121–126, 2022.
- [24] M. I. Dinata and S. Kom, "Interpretasi dan Persamaan Regresi Linier." Accessed: Jun. 21, 2024. [Online]. Available: https://lmsspada.kemdikbud.go.id/pluginfile.php/718616/mod_resource/content/1/bab%201%20Interpretasi%20dan%20Persamaan%20Regresi%20Linier%20%20.pdf
- [25] C. Z. V. Junus, T. Tarno, and P. Kartikasari, "Klasifikasi Menggunakan Metode Support Vector Machine dan Random Forest Untuk Deteksi Awal Resiko Diabetes Melitus," *Jurnal Gaussian*, vol. 11, no. 3, pp. 386–396, Jan. 2023, doi: 10.14710/j.gauss.11.3.386-396.
- [26] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: an Overview," *ArXiv*, Aug. 2020, [Online]. Available: <http://arxiv.org/abs/2008.05756>