

Pengaruh Faktor Genetik Dan Gaya Hidup Terhadap Obesitas Menggunakan Algoritma *K-Nearest Neighbor*

Sherly Meilani, Afrilia Zulian, Imam Sarifhudin, Vlanesa Ginda Ayu Mutia

Sistem Informasi, Universitas Bina Sarana Informatika
Jl. Raya Kaliabang No.8, Perwira, Kec. Bekasi Utara, Kota Bekasi, Jawa Barat, Indonesia

email : sherlymeilani00@gmail.com, afrilia.zulian@gmail.com, imamsarifhudin8723@gmail.com,
vlanesaginda@gmail.com

Submit: 00-00-0000 | Revisi : 00-00-0000 | Terima : 00-00-0000 | Publikasi: 00-00-0000

Abstrak

Obesitas merupakan keadaan dimana seseorang mempunyai berat badan yang melebihi batas normal karena terjadi penumpukan lemak berlebih di dalam tubuh. Obesitas kini menjadi isu kesehatan global yang dipengaruhi oleh berbagai faktor, termasuk faktor genetik serta gaya hidup seperti pola makan, tingkat aktivitas fisik, dan kebiasaan merokok. Penelitian ini bertujuan untuk menganalisis pengaruh faktor genetik dan gaya hidup terhadap obesitas dengan menggunakan metode algoritma *K-Nearest Neighbor* (KNN). Selain itu, aplikasi *Orange* juga digunakan dalam proses klasifikasi setiap individu terhadap potensi obesitas berdasarkan data faktor genetik dan gaya hidup. Penelitian ini diharapkan mampu memberikan wawasan yang lebih mendalam mengenai peran faktor genetik dan gaya hidup dalam memprediksi risiko obesitas pada individu, sekaligus mendukung langkah-langkah pencegahan yang lebih optimal.

Kata Kunci : Obesitas, Genetik, Gaya, Hidup, KNN

Abstract

Obesity is a condition where an individual has a body weight exceeding the normal limit due to excessive fat accumulation in the body. It has become a global health issue influenced by various factors, including genetic factors and lifestyle aspects such as dietary habits, physical activity levels, and smoking. This study aims to analyze the impact of genetic and lifestyle factors on obesity using the K-Nearest Neighbor (KNN) algorithm. Additionally, the Orange application is utilized in the classification process to determine each individual's potential for obesity based on genetic and lifestyle data. This research is expected to provide deeper insights into the role of genetic and lifestyle factors in predicting obesity risk in individuals, as well as support more effective preventive measures.

Keywords : Obesity, Genetic, Life, Style, KNN

1. Pendahuluan

Obesitas kini menjadi isu kesehatan global yang serius, dengan angka prevalensi yang terus bertambah. Kondisi ini dapat dengan mudah diidentifikasi melalui penghitungan Indeks Massa Tubuh (BMI). Individu dengan BMI di atas 25,0 dianggap memiliki berat badan berlebih dan masuk dalam kategori obesitas (Utiarman S. A., 2024). Kondisi ini tidak hanya memengaruhi kualitas hidup individu, tetapi juga meningkatkan risiko berbagai penyakit kronis seperti diabetes melitus, penyakit jantung, dan beberapa jenis kanker. Pemahaman yang mendalam mengenai faktor-faktor yang berkontribusi terhadap obesitas sangat penting untuk pengembangan strategi pencegahan dan pengobatan yang efektif. Selama beberapa dekade terakhir, penelitian telah menunjukkan bahwa obesitas merupakan hasil interaksi kompleks antara faktor genetik dan lingkungan. Faktor genetik dapat mempengaruhi metabolisme tubuh, nafsu makan, dan penyimpanan lemak, sementara gaya hidup yang tidak sehat seperti konsumsi makanan tinggi kalori dan kurangnya aktivitas fisik dapat memicu penumpukan lemak berlebih.

Oleh karena itu, penting untuk mengembangkan metode yang efektif untuk mengklasifikasikan obesitas agar dapat mengambil tindakan pencegahan dan terapi yang tepat. Salah satu dari metode untuk mengklasifikasikan obesitas adalah algoritma *K-Nearest Neighbor* (KNN). Algoritma KNN merupakan metode pembelajaran mesin yang digunakan untuk mengklasifikasikan data berdasarkan hubungannya dengan data pelatihan yang ada. Sebagai bagian dari klasifikasi obesitas, algoritma KNN dapat digunakan untuk



mengklasifikasikan orang menjadi obesitas atau non-obesitas berdasarkan karakteristik dan karakteristik terkait seperti indeks massa tubuh (BMI) dan tinggi badan (Fathoni & Hartanti, 2023).

Tujuan dari penelitian ini adalah mengidentifikasi mutasi genetik yang terkait dengan obesitas, mengevaluasi dampak kebiasaan gaya hidup terhadap risiko obesitas, dan menganalisis pengaruh pola makan, aktivitas fisik, durasi tidur, dan tingkat stres terhadap risiko obesitas, serta mengidentifikasi pola gaya hidup yang paling berisiko, dengan tujuan akhir untuk berkontribusi terhadap pencegahan obesitas dan pengobatan obesitas. Selain itu, penelitian ini juga mengeksplorasi potensi penggunaan algoritma *K-Nearest Neighbor* (KNN) sebagai alat untuk memprediksi risiko obesitas berdasarkan data genetik dan gaya hidup seseorang.

Berdasarkan penelitian sebelumnya dari jurnal Adi Tiyas Ahmad Fathoni dan Dwi Hartanti yang berjudul "Implementasi Algoritma *K-Nearest Neighbor* untuk Klasifikasi Obesitas" memiliki tujuan untuk mengklasifikasikan obesitas dengan menggunakan teknik algoritma *K-Nearest Neighbour* (KNN). Dengan menggunakan kumpulan data yang terdiri dari tinggi dan berat badan, dapat mengklasifikasikan orang menjadi obesitas atau non-obesitas berdasarkan atribut ini. Hasil penelitian ini menunjukkan bahwa model KNN dapat mengklasifikasikan obesitas dengan akurasi yang tinggi, dan dapat membantu mengembangkan metode yang lebih efektif untuk mengklasifikasikan obesitas di masa depan (Fathoni & Hartanti, 2023).

Sedangkan penelitian dari jurnal Koklu, N., dan Sulak, SA yang berjudul "*Using Artificial Intelligence Techniques for the Analysis of Obesity Status According to the Individuals' Social and Physical Activities*" memiliki tujuan untuk mengidentifikasi resiko obesitas dari pihak individu menggunakan kecerdasan buatan. Penelitian ini menggunakan berbagai teknik *machine learning* untuk menentukan status obesitas individu menggunakan kumpulan data yang terdiri dari 14 fitur dan empat kelas. Model *Random Forest* (RF) menghasilkan akurasi klasifikasi tertinggi sebesar 87,82%, Model *Support Vector Machines* (SVM) memiliki akurasi terendah sebesar 74,03, sedangkan model *K-Nearest Neighbor* (KNN) mencapai keberhasilan klasifikasi sebesar 80,62%. Keberhasilan klasifikasi dapat ditingkatkan dengan menggunakan metode pembelajaran mesin yang berbeda (Koklu & Sulak, 2024).

2. Metode

2.1 Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah metode penelitian kuantitatif, karena untuk mengukur keterkaitan antara faktor genetik dan gaya hidup dengan obesitas secara objektif. Metode kuantitatif adalah pendekatan penelitian yang menekankan pada pengumpulan dan analisis data dalam bentuk angka-angka yang dapat diukur dan diuji secara statistik. Pendekatan ini berfokus pada pengujian hipotesis yang spesifik, dilakukan secara sistematis dan terstruktur, dengan penggunaan instrumen yang jelas untuk mengumpulkan data. Dalam penelitian kuantitatif, analisis data dilakukan dengan metode statistik untuk menghasilkan kesimpulan yang objektif dan empiris, yang dapat diuji dan diverifikasi (Waruwu, 2023). Metode penelitian ini menggunakan metode kuantitatif dengan alasan data yang akan diolah yaitu data rasio dan yang menjadi fokus dari penelitian adalah agar mengetahui besarnya pengaruh antar variabel yang dianalisis.

2.2 Metode Algoritma *K-Nearest Neighbor* (KNN)

Metode algoritma yang digunakan dalam analisis pengaruh faktor genetik dan gaya hidup terhadap obesitas ini adalah metode *K-Nearest Neighbor* (KNN). Hal ini dikarenakan metode *K-Nearest Neighbor* (KNN) termasuk metode klasifikasi yang mudah, efisien dan efektif. Algoritma *K-Nearest Neighbor* (KNN) bekerja dengan cara membandingkan data training dan data testing. (Yunus & Pratiwi, 2023)

K-Nearest Neighbor (KNN) merupakan metode *supervised learning* yang digunakan untuk mengklasifikasikan objek baru berdasarkan kedekatannya dengan objek lain. Hasil dari query instance yang baru akan diklasifikasikan berdasarkan kategori yang paling banyak muncul. Dengan kata lain, kelas yang paling sering muncul akan ditetapkan sebagai kelas klasifikasi. Algoritma *K-Nearest Neighbor* (KNN) menerapkan klasifikasi berdasarkan tetangga terdekat sebagai prediksi untuk sampel uji yang baru (Putry & Betha Nurina Sari, 2022).

3. Hasil dan Pembahasan

3.1 Data Collection

Tabel 1. Variabel dataset

No	Item	Penjelasan	Fitur	Jumlah Data
1	Sex	Jenis kelamin individu	1. Laki laki 2. Perempuan	1610
2	Age	Usia individu dalam (tahun)	Nilai dalam bilangan bulat	1610
3	Height	Tinggi badan individu dalam (cm)	Nilai dalam bilangan bulat	1610

			(cm)	
4	Overweight/Obese Families	Riwayat obesitas dalam keluarga	1. Ya 2. Tidak	1610
5	Consumption of Fast Food	Konsumsi makanan cepat saji dalam seminggu	1. Ya 2. Tidak	1610
6	Frequency of Consuming Vegetables	Frekuensi konsumsi sayuran dalam seminggu	1. Jarang 2. Kadang-kadang 3. Selalu	1610
7	Number of Main Meals Daily	Jumlah makanan utama yang dikonsumsi dalam sehari	1. 1-2 2. 3 3. 3+	1610
8	Food Intake Between Meals	Konsumsi Makanan di Antara Waktu Makan Utama (cemilan)	1. Jarang 2. Kadang-kadang 3. Biasanya 4. Selalu	1610
9	Smoking	Kebiasaan Merokok	1. Ya 2. Tidak	1610
10	Liquid Intake Daily	Jumlah cairan yang dikonsumsi dalam sehari (ml)	1. Jumlahnya lebih kecil dari satu liter 2. Dalam kisaran 1- 2 liter 3. Lebih dari 2 liter	1610
11	Calculation Of Calorie Intake	Perhitungan asupan kalori	1. Ya 2. Tidak	1610
12	Physical Exercise	Aktivitas Fisik	1. Tidak ada aktivitas fisik 2. Dalam kisaran 1-2 hari 3. Dalam kisaran 3-4 hari 4. Dalam kisaran 5-6 hari 5. 6+ hari	1610
13	Schedule Dedicated to Technology	Waktu yang Didedikasikan untuk Teknologi	1. Antara 0 dan 2 jam 2. Antara 3 dan 5 jam 3. Melebihi lima jam	1610
14	Type of Transportation Used	Jenis Transportasi yang Digunakan	1. Mobil 2. Motor 3. Sepeda 4. Transportasi umum 5. Jalan kaki	1610
15	Class	Kelas atau kategori yang dimiliki individu	1. Underweight 2. Normal 3. Overweight 4. Obesity	1610

Sumber : Kaggle.com

Dataset yang digunakan berjudul *“Using Artificial Intelligence Techniques for the Analysis of Obesity Status According to the Individuals' Social and Physical Activities”* yang dipublikasikan di Kaggle.com. Pemilihan dataset ini didasarkan pada pertimbangan validitas dan reliabilitas data, yang merupakan syarat mutlak dalam suatu penelitian. Dataset ini telah disusun oleh Koklu, N., dan Sulak, SA, melalui metode pengumpulan data yang telah teruji. Data ini berjumlah 1610 data dengan total 15 variabel yaitu Age, Sex, Height, Overweight or Obese Family, Frequency of Consuming Vegetables, Consumption of Fast Food, Number of Main Meals Daily, Food Intake Between Meals, Smoking, Liquid Intake Daily, Calculation of Calorie Intake, Physical Exercise, Schedule Dedicated to Technology, Type of Transportation Used dan Class.

3.2 Data Selection

Tabel 2. Data Selection

No	Item	Penjelasan	Fitur	Jumlah Data
1	Sex	Jenis kelamin individu	1. Laki laki 2. Perempuan	1610

2	Age	Usia individu dalam (tahun)	Nilai dalam bilangan bulat	1610
3	Height	Tinggi badan individu dalam (cm)	Nilai dalam bilangan bulat (cm)	1610
4	Overweight/Obese Families	Riwayat obesitas dalam keluarga	1. Ya 2. Tidak	1610
5	Consumption of Fast Food	Konsumsi makanan cepat saji dalam seminggu	1. Ya 2. Tidak	1610
6	Number of Main Meals Daily	Jumlah makanan utama yang dikonsumsi dalam sehari	1. 1-2 2. 3 3. 3+	1610
7	Food Intake Between Meals	Konsumsi Makanan di Antara Waktu Makan Utama (cemilan)	1. Jarang 2. Kadang-kadang 3. Biasanya 4. Selalu	1610
8	Smoking	Kebiasaan Merokok	1. Ya 2. Tidak	1610
9	Physical Exercise	Aktivitas Fisik	1. Tidak ada aktivitas fisik 2. Dalam kisaran 1-2 hari 3. Dalam kisaran 3-4 hari 4. Dalam kisaran 5-6 hari 5. 6+ hari	1610
10	Type of Transportation Used	Jenis Transportasi yang Digunakan	1. Mobil 2. Motor 3. Sepeda 4. Transportasi umum 5. Jalan kaki	1610
11	Class	Kelas atau kategori yang dimiliki individu	1. Underweight 2. Normal 3. Overweight 4. Obesity	1610

Sumber : Kaggle.com

Dalam penelitian ini, peneliti menggunakan sampel sebesar 1610 data untuk dianalisis dengan 11 variabel yang secara teoritis memiliki hubungan dengan obesitas, yaitu *Sex*, *Age*, *Height*, *Overweight/Obese Family*, *Consumption of Fast Food*, *Number of Main Meals Daily*, *Food Intake Between Meals*, *Smoking*, *Physical Exercise*, *Type of Transportation Used* dan *Class*. Data-data ini kemudian digunakan untuk membangun model prediksi yang dapat mengklasifikasikan individu ke dalam kategori berat badan: *Underweight*, *Normal*, *Overweight*, atau *Obesity*.

3.3 Data Pre-processing

	Class	Sex	Age	Height	erweight_Obese_Fam	onsumption_of_Fast_Foo	umber_of_Main_Meals_Dail	Food_Intake_Between_Meals	Smoking	Physical_Exercise	pe_of_Transportation
1	Normal	2	18	155	2	2	1	3	2	3	4
2	Normal	2	18	158	2	2	1	1	2	1	3
3	Normal	2	18	159	2	2	1	3	2	2	4
4	Normal	2	18	162	2	2	2	2	2	1	4
5	Normal	2	18	165	2	1	1	3	2	3	2
6	Normal	2	18	176	1	1	1	4	2	4	4
7	Normal	2	19	152	2	2	1	1	2	2	2
8	Normal	2	19	158	2	2	2	4	2	1	3
9	Normal	2	19	159	2	2	2	2	2	2	4
11	Normal	2	19	163	2	2	1	3	2	3	4
12	Normal	2	19	163	2	2	1	1	2	1	4
13	Normal	2	19	165	2	2	1	2	2	2	4
14	Normal	2	19	166	2	1	1	1	2	1	3
15	Normal	2	19	181	1	2	3	4	2	4	4

Gambar 1. Data Pre-processing

Pada tahap data pre-processing dilakukan data cleaning dan penanganan data numerik yang mungkin mengandung nilai yang hilang atau outlier. Data numerik seperti usia (*age*), tinggi badan (*height*), frekuensi makan berat (*number of main meals daily*) dan latihan fisik (*physical exercise*) dapat mengandung nilai yang hilang akibat berbagai alasan. Data yang hilang diisi menggunakan metode yang tepat, sementara outlier

diidentifikasi dan dihapus atau dikoreksi. Langkah-langkah ini bertujuan untuk meningkatkan akurasi dan reliabilitas hasil analisis.

3.4 Evaluasi

Tahapan terakhir dari penelitian ini adalah evaluasi. Pada tahap evaluasi, dilakukan analisis kinerja model *K-Nearest Neighbor (KNN)* menggunakan aplikasi *Orange*. Evaluasi dilaksanakan dengan menggunakan dua metrik yaitu *confusion matrix* dan *scatter plot* sebagai metrik visualisasi data. Sebelum melakukan analisis *confusion matrix* dan *scatter plot*, langkah pertama dalam tahap evaluasi kinerja model *K-Nearest Neighbor* adalah menentukan nilai k . Pada penelitian ini, nilai $k = 2$ dipilih karena memberikan hasil terbaik. Analisis nilai k dilakukan menggunakan metode *cross-validation* dengan fitur *Test & Score* di *Orange*. Percobaan dilakukan dengan berbagai nilai k dari 2 sampai 10. Berdasarkan hasil pengujian, nilai *Classification Accuracy*, *F1 Score*, *Precision*, dan *Recall* tertinggi diperoleh ketika $k = 2$.

Sebagai langkah tambahan untuk memverifikasi hasil, evaluasi juga dapat dilakukan secara manual menggunakan rumus persamaan metrik seperti *accuracy*, *precision*, *recall*, dan *F1 score*. Metrik *accuracy* digunakan untuk mengukur tingkat kedekatan antara hasil prediksi dengan nilai aktual, yang dihitung berdasarkan proporsi data yang diklasifikasikan dengan benar terhadap total data. Adapun persamaan *accuracy* dapat dinyatakan sebagai berikut (Arisandi, Warsito, & Hakim, 2022).

$$Accuracy = \frac{N \text{ benar}}{N} \times 100\% \quad (1)$$

Keterangan:

N = Total data

$N \text{ benar}$ = Jumlah data yang diklasifikasikan benar

Sementara itu, *precision* mengukur seberapa banyak informasi relevan yang berhasil diambil oleh sistem dibandingkan dengan keseluruhan informasi yang diambil, baik yang relevan maupun tidak. Persamaan *precision* diberikan dalam rumus berikut (Arisandi, Warsito, & Hakim, 2022).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Keterangan:

TP (*True Positives*) = Jumlah data positif yang berhasil diklasifikasikan dengan tepat oleh sistem

FP (*False Positives*) = Jumlah data negatif yang keliru diklasifikasikan sebagai positif

Recall digunakan untuk mengevaluasi kemampuan sistem dalam mengambil seluruh informasi relevan yang tersedia, termasuk yang tidak terambil. Rumus untuk *recall* ditunjukkan sebagai berikut (Arisandi, Warsito, & Hakim, 2022).

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Keterangan:

TP (*True Positives*) = Jumlah data positif yang diklasifikasikan dengan benar oleh sistem

FN (*False Negatives*) = Jumlah data positif yang diklasifikasikan secara salah sebagai negatif

Terakhir, *F-measure*, atau yang dikenal sebagai *F1 score*, merupakan rata-rata harmonis antara *precision* dan *recall*, yang bertujuan untuk memberikan gambaran keseimbangan antara kedua metrik tersebut. Persamaan *F1 score* dirumuskan sebagai berikut (Arisandi, Warsito, & Hakim, 2022).

$$F1 \text{ Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Evaluation results for target (None, show average over classes) ▾						
Model	AUC	CA	F1	Prec	Recall	MCC
kNN (1)	0.877	0.795	0.793	0.793	0.795	0.689

Gambar 2. Evaluasi test and score

Data tersebut menunjukkan hasil evaluasi kinerja model *K-Nearest Neighbor* (KNN) berdasarkan berbagai metrik, yang membantu dalam memahami seberapa baik model dalam mengklasifikasikan data. *AUC* (*Area Under the Curve*) merupakan metrik yang menilai kemampuan model untuk membedakan kelas positif dan negatif, dengan nilai berkisar antara 0 hingga 1. Semakin mendekati 1, maka semakin baik. Nilai 0.877 menunjukkan bahwa model memiliki performa yang cukup baik dalam membedakan kelas. *CA* (*Classification Accuracy*) merupakan persentase prediksi benar, dengan nilai 0.795 berarti 79.5% prediksi model akurat. *F1 Score* merupakan rata-rata harmonis antara *Precision* dan *Recall*, nilai 0.793 menunjukkan keseimbangan yang cukup baik.

Precision adalah proporsi prediksi positif yang benar, nilai 0.793 menunjukkan akurasi 79.3% dalam prediksi positif. *Recall* atau *Sensitivity* adalah proporsi data kelas positif yang benar-benar teridentifikasi oleh model. Nilai 0.795 berarti model berhasil mengidentifikasi 79.5% dari total data kelas positif yang ada. *MCC* (*Matthews Correlation Coefficient*) adalah metrik yang mengukur kekuatan hubungan antara prediksi model dan label actual yang mempertimbangkan ketidakseimbangan kelas. Nilai *MCC* berkisar antara -1 dan 1, di mana 1 menunjukkan prediksi sempurna, 0 menunjukkan performa acak, dan -1 menunjukkan prediksi yang benar-benar salah. Nilai 0.689 berarti hubungan prediksi dan label actual cukup baik. Secara keseluruhan, model *K-Nearest Neighbor* (KNN) ini memiliki performa yang cukup baik, terutama dengan nilai *AUC* yang tinggi.

a. Confusion matrix

Confusion matrix digunakan sebagai alat untuk menilai kinerja model klasifikasi. Tabel ini menunjukkan perbandingan antara prediksi yang dibuat oleh model dan label sebenarnya dari data. Dengan menggunakan *confusion matrix*, bisa didapatkan informasi yang jelas tentang seberapa baik model dalam mengklasifikasikan data ke dalam kelas yang benar.

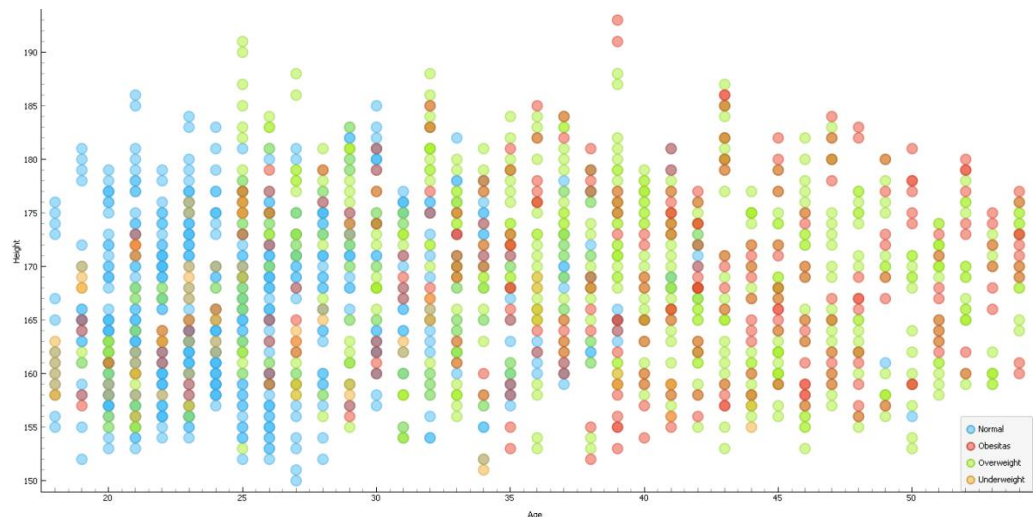
		Predicted				
		Normal	Obesitas	Overweight	Underweight	Σ
Actual	Normal	588	9	52	9	658
	Obesitas	20	201	63	3	287
	Overweight	72	66	452	2	592
	Underweight	27	3	4	39	73
Σ		707	279	571	53	1610

Gambar 3. Confusion Matrix

Baris pada tabel tersebut menunjukan kelas *actual* atau nilai sebenarnya sedangkan kolom menunjukan kelas yang diprediksi oleh model. Prediksi yang benar ditandai dengan warna biru sementara warna merah menunjukan bahwa model salah memprediksi kelas. Gambar di atas menginterpretasi bahwa ada 588 individu yang sebenarnya memiliki berat badan normal dan juga diprediksi sebagai berat badan normal oleh model. Data tersebut termasuk ke dalam prediksi yang benar. Selain itu, prediksi yang salah ditunjukkan pada data dimana 72 individu yang sebenarnya kelebihan berat badan (*overweight*) namun justru diprediksi sebagai berat badan normal oleh model.

b. Scatter plot

Metrik *Scatter plot* digunakan dalam proses visualisasi data hasil *K-Nearest Neighbor* (KNN). Grafik ini berfungsi untuk menggambarkan hubungan antar variabel yang dianalisis. Hal ini dilakukan untuk mendapatkan informasi analisis lebih lanjut sehingga data dengan mudah disimpulkan.



Gambar 4. Scatter Plot Pengaruh Usia dan Tinggi Badan terhadap Klasifikasi Berat Badan

Gambar di atas adalah *scatter plot* yang menampilkan hubungan antara variabel usia, tinggi dan klasifikasi berat badan. Sumbu x mewakili variabel usia (*age*) sementara sumbu y mewakili tinggi badan (*height*). Lingkaran berwarna merah menggambarkan individu dengan obesitas, hijau untuk *overweight*, biru untuk normal, dan oranye untuk *underweight*.

Grafik menunjukkan bahwa individu dengan usia di atas 35 tahun memiliki proporsi yang signifikan dalam kategori berat badan *overweight* atau bahkan obesitas. Temuan ini mengindikasikan bahwa orang yang lebih tua, terutama yang memiliki tinggi badan di atas rata-rata, cenderung memiliki risiko lebih besar untuk mengalami *overweight* atau obesitas. Dari *scatter plot* ini terlihat adanya korelasi positif antara bertambahnya usia dan peningkatan risiko memiliki berat badan *overweight* atau obesitas. Temuan ini merupakan *insight* penting yang perlu diperhatikan, terutama dalam perancangan program pencegahan obesitas untuk kelompok usia tersebut.

4. Kesimpulan

Penelitian ini mengidentifikasi faktor genetik dan gaya hidup yang signifikan berkontribusi pada risiko obesitas, sejalan dengan tujuan awal penelitian. Penerapan algoritma *K-Nearest Neighbor* menunjukkan potensi yang menjanjikan dalam memprediksi risiko obesitas, sebagaimana dibuktikan oleh hasil klasifikasi yang membedakan antara individu dengan berat badan normal, kekurangan berat badan, kelebihan berat badan, dan obesitas. Hasil dari penelitian ini adalah sebagai berikut.

- Hasil CA (*Classification Accuracy*) 79.5%, AUC (*Area Under the Curve*) 0.877, F1 Score 0.793, Precision 0.793, Recall 0.795, MCC (*Matthews Correlation Coefficient*) 0.689. Secara keseluruhan, model *K-Nearest Neighbor (KNN)* ini memiliki performa yang cukup baik.
- Hasil penelitian menunjukkan bahwa distribusi kategori berat badan sebagai berikut: 73 individu dikategorikan sebagai *underweight*, 658 individu sebagai normal, 592 individu sebagai *overweight*, dan 287 individu sebagai obesitas. Temuan ini mengindikasikan bahwa mayoritas partisipan dalam penelitian ini memiliki berat badan normal atau kelebihan berat badan.

Penelitian ini mendukung pengembangan intervensi yang lebih tepat sasaran dan efektif dalam pencegahan serta penanganan obesitas. Meski demikian, perlu diakui bahwa terdapat beberapa keterbatasan dalam penelitian ini, diantaranya sampelnya sedikit dan hasilnya mungkin tidak bisa dianggap mewakili populasi yang lebih luas. Untuk penelitian selanjutnya disarankan dapat menjelajahi lebih dalam mekanisme genetik yang mendasari obesitas, mengevaluasi efektivitas intervensi berbasis data yang lebih luas, dan mempertimbangkan interaksi antara faktor genetik serta lingkungan yang lebih kompleks. Dengan demikian, penelitian ini memberikan kontribusi penting dalam upaya global untuk mengatasi masalah obesitas.

Referensi

- Arisandi, R. R., Warsito, B., & Hakim, A. R. (2022). APLIKASI NAÏVE BAYES CLASSIFIER (NBC) PADA KLASIFIKASI STATUS GIZI BALITA STUNTING DENGAN PENGUJIAN K-FOLD CROSS VALIDATION. *JURNAL GAUSSIAN*, 130 - 139.
- Dewia, A. M., & Dwidasmara, I. B. (2020). Implementation Of The K-Nearest Neighbor (KNN) Algorithm For Classification Of Obesity Levels. *Jurnal Elektronik Ilmu Komputer Udayana*, 277-284.
- Fathoni, A. T., & Hartanti, D. (2023). Implementasi Algoritma K-Nearest Neighbor untuk Klasifikasi Obesitas. *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, 634-639.
- Hardwis, S., & Jajat. (2024). Analisis Resiko Obesitas Berdasarkan Aktivitas Fisik: Implementasi Metode Artificial Intelligence Machine Learning. *Jurnal Keolahragaan*, 29-36.
- Koklu, N., & Sulak, S. (2024). Using Artificial Intelligence Techniques for the Analysis of Obesity Status According to the Individuals' Social and Physical Activities. *Sinop Üniversitesi Fen Bilimleri Dergisi*, 217-239.
- Putry, N. M., & Betha Nurina Sari, M. (2022). KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELITUS. *Evolusi: Jurnal Sains dan Manajemen*, 45-57.
- Setiyani, L., Indahsari, A. N., & Roestam, R. (2023). Analisis Prediksi Level Obesitas Menggunakan Perbandingan Algoritma Machine Learning dan Deep Learning. *JTERA (Jurnal Teknologi Rekayasa)*, 139-146.
- Utiahman, S. A., & Pratama, A. M. (2024). Analisis Perbandingan KNN, SVM, Decision Tree dan Regresi Logistik Untuk Klasifikasi Obesitas Multi Kelas. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 3137-3146.
- Waruwu, M. (2023). Pendekatan Penelitian Pendidikan: Metode Penelitian Kualitatif, Metode Penelitian Kuantitatif dan Metode Penelitian Kombinasi (Mixed Method). *Jurnal Pendidikan Tambusai*, 2896-2910.
- Yunus, M., & Pratiwi, N. K. (2023). Prediksi Status Gizi Balita Dengan Algoritma K-Nearest Neighbor (KNN) di Puskesmas Cakranegara. *JTIM : Jurnal Teknologi Informasi dan Multimedia*, 221-231.