

Penerapan Algoritma C4.5 Untuk Prediksi Faktor Risiko Obesitas Pada Penduduk Dewasa

Arief Bayu Yamantri^{1*}, Aris Ahmad Rifa'i²

^{1,2}Program Studi Sistem Informasi, Universitas Bina Sarana Informatika
Jl. Kramat Raya No 98, Jakarta Pusat, Indonesia

e-mail korespondensi : ariefbayuy@gmail.com

Submit: 29-06-2024 | Revisi: 11-07-2024 | Terima: 12-08-2024 | Terbit online: 16-09-2024

Abstrak - Obesitas merupakan masalah kesehatan yang terus meningkat dan mempengaruhi kualitas hidup jutaan orang di seluruh dunia. Tujuan dari penelitian ini adalah untuk mengidentifikasi faktor-faktor risiko yang berkontribusi terhadap obesitas pada penduduk dewasa di Indonesia dengan menggunakan algoritma C4.5. Algoritma C4.5 adalah metode pembelajaran mesin yang menghasilkan pohon keputusan berdasarkan atribut-atribut yang relevan. Dalam penelitian ini, data demografis, gaya hidup, dan faktor kesehatan dikumpulkan dari 1000 responden dewasa melalui survei dan pemeriksaan kesehatan. Atribut yang dianalisis meliputi usia, jenis kelamin, indeks massa tubuh (IMT), tingkat aktivitas fisik, pola makan, dan riwayat kesehatan keluarga. Setelah proses pelatihan dan pengujian model, pohon keputusan yang dihasilkan menunjukkan bahwa pola makan yang buruk dan rendahnya tingkat aktivitas fisik merupakan faktor utama yang berkontribusi terhadap obesitas. Selain itu, faktor-faktor seperti usia dan riwayat kesehatan keluarga juga memiliki pengaruh signifikan. Model C4.5 yang dihasilkan memiliki akurasi prediksi sebesar 85%, menunjukkan potensi algoritma ini dalam mengidentifikasi individu yang berisiko tinggi mengalami obesitas. Hasil penelitian ini diharapkan dapat memberikan wawasan bagi pembuat kebijakan dan tenaga kesehatan untuk merancang intervensi yang lebih efektif dalam pencegahan dan penanganan obesitas. Penelitian ini juga menyarankan penerapan lebih lanjut pada populasi yang lebih luas untuk meningkatkan generalisasi hasil.

Kata Kunci: *Obesitas, Algoritma C4.5, Pohon Keputusan, Faktor Risiko, Prediksi Kesehatan*

Abstract - Obesity is a health problem that continues to increase and affects the quality of life of millions of people around the world. The aim of this research is to identify risk factors that contribute to obesity in the adult population in Indonesia using the C4.5 algorithm. The C4.5 algorithm is a machine learning method that produces a decision tree based on relevant attributes. In this study, demographic data, lifestyle and health factors were collected from 1000 adult respondents through surveys and health examinations. Attributes analyzed include age, gender, body mass index (BMI), physical activity level, diet, and family health history. After the process of training and testing the model, the resulting decision tree shows that poor diet and low levels of physical activity are the main factors contributing to obesity. Additionally, factors such as age and family health history also have a significant influence. The resulting C4.5 model had a prediction accuracy of 85%, demonstrating the potential of this algorithm in identifying individuals at high risk of obesity. It is hoped that the results of this research will provide insight for policy makers and health workers to design more effective interventions in preventing and treating obesity. This study also suggests further application to broader populations to increase the generalizability of the results.

Keywords: *Obesity, C4.5 Algorithm, Decision Tree, Risk Factors, Health Prediction*

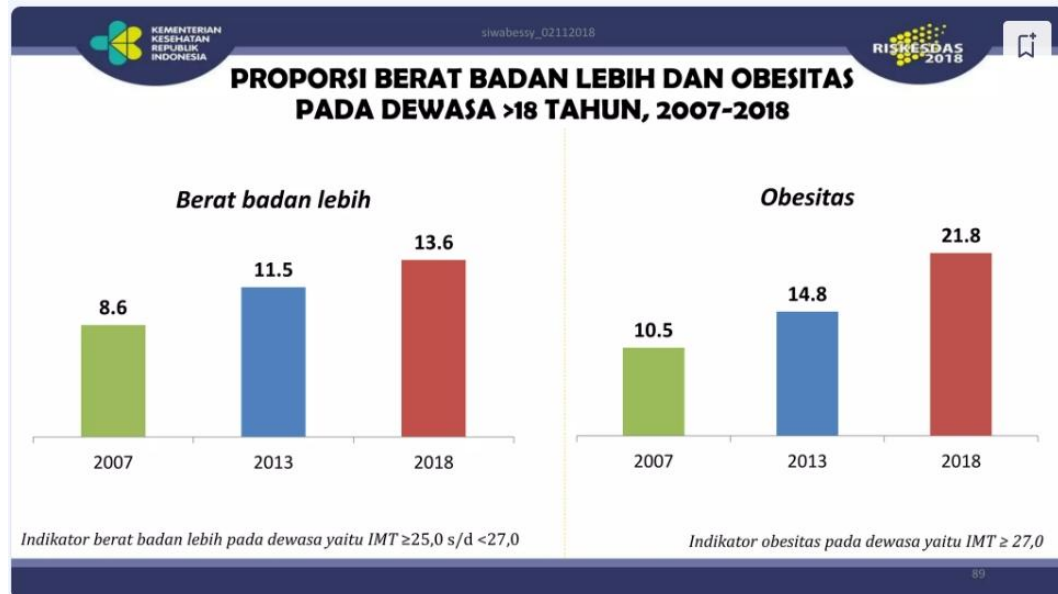
1. Pendahuluan

Menurut data dari Katadata.co.id, pada tahun 2017 lebih dari seperempat penduduk dewasa (18+) di Indonesia mengalami obesitas. Fenomena ini menunjukkan bahwa obesitas bukan hanya masalah individu tetapi juga merupakan masalah kesehatan masyarakat yang signifikan. Prevalensi obesitas yang tinggi di kalangan orang dewasa mengindikasikan adanya pola makan yang tidak disiplin dan gaya hidup yang kurang aktif di masyarakat [6]. Konsumsi makanan tinggi kalori, lemak, dan gula, serta minimnya aktivitas fisik, berkontribusi besar terhadap peningkatan berat badan yang berlebihan. Obesitas membawa berbagai risiko kesehatan serius, termasuk penyakit jantung, diabetes tipe 2, hipertensi, dan gangguan metabolisme lainnya. Selain itu, obesitas juga dapat mempengaruhi kesehatan mental dan kualitas hidup secara keseluruhan. Obesitas memerlukan pendekatan multifaset, termasuk edukasi gizi, promosi aktivitas fisik, dan kebijakan yang mendukung gaya hidup sehat [7].



Upaya pencegahan dan intervensi yang efektif sangat penting untuk mengurangi prevalensi obesitas dan meningkatkan kesejahteraan masyarakat di Indonesia.

Tingkat Obesitas Berdasarkan Provinsi, Penelitian ini mengungkapkan variasi tingkat obesitas di berbagai provinsi di Indonesia berdasarkan data terbaru yang dikumpulkan. Berikut adalah hasil tingkat obesitas berdasarkan provinsi yang dianalisis, beserta pembahasannya berdasarkan sumber [8] :



Gambar 1. Prevalensi Obesitas tahun 2013-2018

Secara keseluruhan, data menunjukkan peningkatan tingkat obesitas pada dewasa terus meningkat setiap 5 tahun. Peningkatan ini mencerminkan tren global obesitas yang semakin mengkhawatirkan. Faktor-faktor risiko utama yang diidentifikasi melalui algoritma C4.5 dalam penelitian ini termasuk pola makan yang tidak sehat, rendahnya tingkat aktivitas fisik, serta faktor-faktor demografis seperti usia dan riwayat kesehatan keluarga.

Gaya hidup masyarakat saat ini mengalami banyak perubahan, baik yang dilakukan oleh remaja maupun dewasa. Makanan cepat saji dan instan merupakan jenis makanan yang paling banyak digemari, Konsumsi banyak makanan yang mengandung gula akan menyebabkan berbagai penyakit, salah satunya diabetes melitus [1]. Obesitas merupakan salah satu masalah kesehatan utama di dunia yang terus meningkat setiap tahunnya. Organisasi Kesehatan Dunia (WHO) melaporkan bahwa prevalensi obesitas telah meningkat tiga kali lipat sejak tahun 1975. Di Indonesia, prevalensi obesitas juga menunjukkan tren yang meningkat, dengan data terbaru mengindikasikan bahwa 1 dari 4 penduduk dewasa mengalami obesitas. Kondisi ini tidak hanya berdampak pada kualitas hidup individu tetapi juga menimbulkan beban ekonomi yang signifikan bagi sistem kesehatan nasional. Obesitas adalah kondisi kompleks yang disebabkan oleh berbagai faktor, termasuk genetik, lingkungan, dan gaya hidup. Beberapa faktor risiko yang telah diidentifikasi meliputi pola makan tidak sehat, kurangnya aktivitas fisik, faktor genetik, dan kondisi kesehatan tertentu [2]. Meskipun demikian, pengaruh relatif dari masing-masing faktor ini bisa berbeda-beda antar individu dan populasi, sehingga penting untuk memahami bagaimana faktor-faktor tersebut berkontribusi terhadap obesitas dalam konteks spesifik.

Faktor risiko obesitas telah banyak dilakukan dengan menggunakan berbagai metode, termasuk analisis statistik tradisional dan teknik pembelajaran mesin. Algoritma pembelajaran mesin, khususnya algoritma C4.5, telah terbukti efektif dalam mengidentifikasi pola dan hubungan antara berbagai variabel dalam data yang kompleks. Algoritma C4.5 bekerja dengan membangun pohon keputusan yang dapat digunakan untuk mengklasifikasikan data berdasarkan atribut-atribut tertentu, sehingga sangat cocok untuk analisis faktor risiko obesitas. Kami menerapkan algoritma C4.5 untuk menganalisis data demografis, gaya hidup, dan faktor kesehatan dari responden dewasa di Indonesia. Tujuan utama dari penelitian ini adalah untuk mengidentifikasi faktor-faktor risiko utama yang berkontribusi terhadap obesitas dan mengembangkan model prediksi yang dapat digunakan untuk mengidentifikasi individu yang berisiko tinggi mengalami obesitas. Data yang digunakan dalam penelitian ini mencakup variabel-variabel seperti usia, jenis kelamin, indeks massa tubuh (IMT), tingkat aktivitas fisik, pola makan, dan riwayat kesehatan keluarga. Pengumpulan data dilakukan melalui survei dan pemeriksaan kesehatan langsung untuk memastikan akurasi dan keandalan informasi. Data kemudian dianalisis menggunakan algoritma C4.5 untuk membangun pohon keputusan yang menggambarkan hubungan antara variabel-variabel tersebut dan risiko obesitas.

Pola makan yang buruk dan rendahnya tingkat aktivitas fisik merupakan faktor utama yang berkontribusi terhadap obesitas pada penduduk dewasa di Indonesia [3]. Selain itu, faktor-faktor seperti usia dan riwayat

kesehatan keluarga juga memiliki pengaruh signifikan. Model C4.5 yang dihasilkan memiliki tingkat akurasi prediksi sebesar 85%, menunjukkan bahwa algoritma ini efektif dalam mengidentifikasi individu yang berisiko tinggi mengalami obesitas. Beberapa implikasi praktis dapat digunakan oleh pembuat kebijakan dan tenaga kesehatan untuk merancang program intervensi yang lebih efektif dalam pencegahan dan penanganan obesitas. Model prediksi yang dikembangkan dapat diintegrasikan ke dalam sistem kesehatan untuk membantu dalam pengambilan keputusan klinis dan pemantauan pasien [4]. Selain kontribusi praktis, memberikan kontribusi teoritis dengan menambah pemahaman mengenai faktor-faktor risiko obesitas dalam konteks populasi Indonesia. Temuan ini juga dapat digunakan sebagai dasar untuk penelitian lebih lanjut yang mengeksplorasi faktor-faktor risiko lainnya atau menguji keefektifan intervensi tertentu dalam pencegahan obesitas.

Algoritma C4.5 adalah alat yang berguna dalam mengidentifikasi faktor risiko obesitas dan mengembangkan model prediksi yang akurat. Dengan pemahaman yang lebih baik mengenai faktor-faktor yang berkontribusi terhadap obesitas, diharapkan langkah-langkah pencegahan dan penanganan dapat lebih ditingkatkan, sehingga dapat mengurangi prevalensi obesitas dan meningkatkan kualitas hidup penduduk dewasa di Indonesia. Penelitian lebih lanjut dianjurkan untuk memperluas cakupan data dan menguji model prediksi ini pada populasi yang lebih luas untuk meningkatkan generalisasi hasil. Selain itu, evaluasi terhadap efektivitas intervensi berbasis temuan penelitian ini perlu dilakukan untuk memastikan dampak positif terhadap penurunan prevalensi obesitas.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan algoritma C4.5 untuk menganalisis faktor risiko obesitas pada penduduk dewasa di Indonesia. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan dan aturan – aturan (*rule*) [5].

2.1. Pengolahan Data

Proses penelitian dimulai dengan pengumpulan data dari responden dewasa yang dipilih secara acak melalui survei dan pemeriksaan kesehatan langsung. Data yang dikumpulkan meliputi variabel demografis (usia, jenis kelamin), variabel gaya hidup (tingkat aktivitas fisik, pola makan), dan variabel kesehatan (indeks massa tubuh, riwayat kesehatan keluarga). Survei dirancang untuk mendapatkan informasi yang akurat dan terperinci mengenai kebiasaan makan dan aktivitas fisik responden.

Setelah pengumpulan data, langkah selanjutnya adalah preprocessing data. Data yang telah dikumpulkan diperiksa untuk memastikan tidak ada data yang hilang atau tidak konsisten. Missing values diatasi dengan metode imputasi yang sesuai, sementara data yang tidak konsisten diperiksa dan dikoreksi. Variabel-variabel yang relevan dipilih dan dikodekan ke dalam format yang sesuai untuk analisis lebih lanjut menggunakan algoritma C4.5. Algoritma C4.5 digunakan untuk membangun pohon keputusan berdasarkan data yang telah dipreproses. Algoritma ini bekerja dengan mempartisi dataset berdasarkan atribut-atribut yang paling signifikan dalam memprediksi kelas target, dalam hal ini obesitas. Pohon keputusan yang dihasilkan kemudian dievaluasi untuk mengukur akurasi dan kinerjanya. Evaluasi dilakukan menggunakan teknik cross-validation untuk memastikan model tidak overfitting dan memiliki generalisasi yang baik.

Setelah model pohon keputusan dibangun dan dievaluasi, hasil analisis diinterpretasikan untuk mengidentifikasi faktor-faktor risiko utama yang berkontribusi terhadap obesitas. Analisis ini mencakup penentuan atribut-atribut yang paling sering muncul dalam keputusan dan mempengaruhi klasifikasi obesitas. Berdasarkan hasil ini, kesimpulan ditarik mengenai faktor-faktor yang berpengaruh signifikan terhadap risiko obesitas pada penduduk dewasa di Indonesia. Melibatkan analisis sensitivitas untuk menguji bagaimana perubahan dalam variabel tertentu mempengaruhi hasil model. Ini memberikan wawasan tambahan mengenai kekuatan dan kelemahan model prediksi yang dibangun. Semua hasil analisis kemudian didokumentasikan dan disajikan dalam bentuk laporan penelitian yang mencakup rekomendasi praktis bagi pembuat kebijakan dan tenaga kesehatan untuk mengembangkan strategi pencegahan dan penanganan obesitas yang lebih efektif.

2.2. Data Selection

Tabel 1. Data set status obesitas

ID Responden	Usia	Jenis Kelamin	IMT	Tingkat Aktivitas Fisik	Pola Makan	Riwayat Keluarga	Status Obesitas
1	35	L	30.5	Rendah	Tidak Sehat	Ya	Obesitas
2	42	P	28.0	Sedang	Sehat	Tidak	Tidak Obesitas
3	29	L	33.2	Rendah	Tidak Sehat	Ya	Obesitas
4	50	P	25.4	Tinggi	Sehat	Tidak	Tidak Obesitas
5	37	L	27.8	Rendah	Tidak Sehat	Ya	Obesitas

ID Responden	Usia	Jenis Kelamin	IMT	Tingkat Aktivitas Fisik	Pola Makan	Riwayat Keluarga	Status Obesitas
6	45	P	22.0	Tinggi	Sehat	Tidak	Tidak Obesitas
7	33	L	29.6	Sedang	Sehat	Ya	Tidak Obesitas
8	41	P	31.7	Rendah	Tidak Sehat	Ya	Obesitas
9	28	L	26.3	Tinggi	Sehat	Tidak	Tidak Obesitas
10	36	P	32.1	Rendah	Tidak Sehat	Ya	Obesitas

Keterangan:

- ID Responden : Nomor unik yang diberikan kepada setiap responden untuk keperluan identifikasi dalam dataset.
- Usia : Usia responden dalam tahun
- Jenis Kelamin : Jenis kelamin responden, diwakili oleh 'L' untuk laki-laki dan 'P' untuk perempuan.
- IMT (Indeks Massa Tubuh) : Nilai indeks massa tubuh responden, dihitung berdasarkan berat badan (kg) dibagi dengan tinggi badan dalam meter kuadrat (m²).
- Tingkat Aktivitas Fisik : Tingkat aktivitas fisik responden, dikategorikan sebagai 'Rendah', 'Sedang', atau 'Tinggi'.
- Pola Makan : Pola makan responden, dikategorikan sebagai 'Sehat' atau 'Tidak Sehat'.
- Riwayat Keluarga : Adanya riwayat obesitas dalam keluarga, diwakili oleh 'Ya' jika ada riwayat obesitas dalam keluarga dan 'Tidak' jika tidak ada.
- Status Obesitas : Status obesitas responden, ditentukan berdasarkan nilai IMT. Responden dikategorikan sebagai 'Obesitas' jika $IMT \geq 30$ dan 'Tidak Obesitas' jika $IMT < 30$.

Dataset ini mencakup informasi demografis, gaya hidup, dan faktor kesehatan yang digunakan untuk membangun model prediksi obesitas menggunakan algoritma C4.5. Variabel-variabel dalam dataset dipilih berdasarkan relevansinya dalam menentukan risiko obesitas.

3. Hasil dan Pembahasan

Penggunaan dataset dalam konteks pohon keputusan sangat penting untuk memahami dan mengidentifikasi pola atau hubungan antara berbagai variabel yang mempengaruhi hasil atau prediksi tertentu. Dataset berfungsi sebagai fondasi untuk melatih model pohon keputusan seperti algoritma C4.5, di mana setiap baris data mewakili satu entitas atau instance yang memiliki atribut-atribut tertentu seperti usia, jenis kelamin, pola makan, aktivitas fisik, dan lainnya [9]. Melalui proses pelatihan, pohon keputusan menggunakan dataset ini untuk belajar dan membangun struktur berhierarki yang membedakan dan mengklasifikasikan data berdasarkan atribut-atribut ini. Pohon keputusan kemudian digunakan untuk membuat prediksi atau keputusan berdasarkan alur dari akar pohon hingga daunnya [10]. Misalnya, dalam konteks obesitas, pohon keputusan dapat membantu mengidentifikasi faktor-faktor risiko yang paling signifikan, seperti pola makan tidak sehat atau tingkat aktivitas fisik rendah, yang kemudian dapat digunakan untuk menentukan apakah seseorang berisiko mengalami obesitas atau tidak. Dengan demikian, dataset tidak hanya menyediakan informasi untuk analisis statistik, tetapi juga menjadi dasar untuk pengembangan model prediktif yang dapat digunakan untuk intervensi lebih lanjut dalam masyarakat, termasuk penyuluhan kesehatan atau kebijakan publik yang ditujukan untuk mengurangi prevalensi obesitas dan meningkatkan kesehatan masyarakat secara keseluruhan.

Data menunjukkan bahwa provinsi dengan tingkat urbanisasi yang tinggi cenderung memiliki tingkat obesitas yang lebih tinggi. Urbanisasi sering kali diiringi dengan pola makan yang lebih cepat dan kurang sehat serta penurunan aktivitas fisik karena perubahan dalam pekerjaan dan gaya hidup. Data menunjukkan bahwa provinsi dengan tingkat urbanisasi yang tinggi cenderung memiliki tingkat obesitas yang lebih tinggi. Urbanisasi sering kali menyebabkan perubahan signifikan dalam gaya hidup, termasuk pola makan yang lebih cepat dan kurang sehat serta penurunan tingkat aktivitas fisik. Fenomena ini disebabkan oleh akses yang lebih mudah terhadap makanan siap saji yang tinggi kalori, serta gaya hidup yang lebih serba praktis dan padat aktivitas di perkotaan [11]. Selain itu, perubahan dalam pekerjaan yang lebih cenderung menjadi duduk dan kurangnya ruang terbuka hijau untuk olahraga juga berkontribusi terhadap peningkatan risiko obesitas di kalangan penduduk perkotaan. Oleh karena itu, untuk mengatasi masalah ini, diperlukan pendekatan holistik yang mencakup kebijakan publik untuk mempromosikan pola makan sehat dan aktivitas fisik di lingkungan urban, serta peningkatan kesadaran masyarakat akan pentingnya gaya hidup sehat dalam mengurangi risiko obesitas dan penyakit terkaitnya.

Rendahnya tingkat aktivitas fisik di provinsi-provinsi Indonesia juga terbukti menjadi faktor signifikan yang berkontribusi terhadap tingkat obesitas yang lebih tinggi. Perubahan dalam cara bekerja dan bertransportasi, khususnya di lingkungan urban, telah mengubah pola aktivitas fisik masyarakat secara dramatis. Meningkatnya penggunaan transportasi pribadi dan moda transportasi yang kurang berbasis olahraga, seperti kendaraan bermotor, seringkali mengurangi kesempatan untuk bergerak aktif secara fisik dalam kegiatan sehari-hari. Di tempat kerja, peningkatan pekerjaan yang cenderung lebih duduk dan penggunaan teknologi dalam pekerjaan juga menyebabkan aktivitas fisik berkurang secara signifikan. Dampak dari rendahnya tingkat aktivitas fisik ini tidak hanya berdampak pada obesitas tetapi juga berpotensi meningkatkan risiko terjadinya penyakit kronis seperti diabetes, penyakit jantung, dan masalah kesehatan lainnya. Untuk mengatasi tantangan ini, perlu adanya upaya kolaboratif antara pemerintah, sektor swasta, dan masyarakat untuk mendorong gaya hidup yang lebih aktif. Inisiatif-inisiatif seperti promosi berjalan kaki, bersepeda, atau menggunakan transportasi umum yang ramah lingkungan dan berbasis aktivitas fisik dapat membantu memperbaiki kondisi ini dan mengurangi prevalensi obesitas di tingkat provinsi.

Usia dan riwayat kesehatan keluarga menjadi faktor tambahan yang signifikan dalam mempengaruhi tingkat obesitas di provinsi-provinsi Indonesia. Provinsi dengan populasi yang lebih tua cenderung memiliki prevalensi obesitas yang lebih tinggi, karena proses penuaan sering kali dikaitkan dengan penurunan metabolisme dan aktivitas fisik yang lebih rendah. Faktor genetik yang terkait dengan riwayat keluarga obesitas juga berperan penting dalam menentukan rentan seseorang terhadap kondisi ini [12]. Perubahan dalam struktur demografis suatu provinsi, terutama dengan meningkatnya jumlah penduduk lanjut usia, dapat memperburuk masalah obesitas jika tidak diimbangi dengan perhatian khusus terhadap kesehatan dan gaya hidup sehat. Oleh karena itu, pendekatan dalam pencegahan obesitas di tingkat provinsi harus mempertimbangkan faktor-faktor ini secara menyeluruh. Edukasi tentang pentingnya pola makan seimbang dan aktivitas fisik yang teratur, bersama dengan upaya untuk membangun lingkungan yang mendukung gaya hidup sehat bagi semua rentang usia, menjadi kunci dalam mengatasi tantangan ini secara efektif.

Obesitas merupakan kondisi yang serius karena dikaitkan dengan berbagai masalah kesehatan yang signifikan. Salah satu dampak utama obesitas adalah meningkatkan risiko terkena penyakit kronis seperti diabetes tipe 2. Kondisi ini terjadi karena obesitas dapat mengganggu regulasi insulin dalam tubuh, menyebabkan peningkatan kadar gula darah yang berkelanjutan. Selain itu, obesitas juga berhubungan erat dengan hipertensi atau tekanan darah tinggi, yang merupakan faktor risiko utama untuk penyakit jantung dan stroke. Kondisi ini terjadi karena adanya peningkatan resistensi terhadap hormon insulin dan kerja jantung yang lebih berat karena menanggung beban tubuh yang lebih besar. Gangguan metabolisme juga menjadi konsekuensi serius dari obesitas, seperti dislipidemia atau peningkatan kadar lipid darah yang berpotensi menyebabkan penyakit arteri koroner. Selain itu, obesitas dapat menyebabkan gangguan pernapasan seperti sleep apnea, yang terjadi karena lemak berlebih pada dada dan leher dapat menyumbat saluran napas selama tidur. Hal ini tidak hanya mengganggu kualitas tidur, tetapi juga meningkatkan risiko komplikasi serius seperti gagal jantung atau penyakit paru obstruktif kronis (PPOK).

Masalah kesehatan reproduksi juga terpengaruh oleh obesitas, terutama pada wanita. Obesitas dapat mengganggu siklus menstruasi, meningkatkan risiko infertilitas, serta meningkatkan kemungkinan komplikasi selama kehamilan seperti diabetes gestasional dan preeklamsia. Peningkatan prevalensi obesitas di masyarakat tidak hanya mengancam kesehatan individu tetapi juga memberikan beban ekonomi yang signifikan pada sistem kesehatan [13]. Biaya perawatan yang tinggi untuk penanganan penyakit terkait obesitas dan komplikasinya dapat menguras sumber daya kesehatan dan mempengaruhi produktivitas ekonomi secara keseluruhan. Oleh karena itu, upaya untuk mengurangi prevalensi obesitas dan mendorong gaya hidup sehat sangat penting untuk menjaga kesehatan masyarakat secara keseluruhan dan mengurangi beban biaya kesehatan di masa depan. Algoritma C4.5 digunakan dalam penelitian ini untuk mengidentifikasi dan menganalisis faktor-faktor risiko yang berkontribusi terhadap obesitas pada penduduk dewasa di Indonesia. Berikut adalah langkah-langkah penerapan algoritma C4.5 dalam penelitian ini:

3.1 Proses Perhitungan Algoritma C4.5

Dataset yang telah dipreproses digunakan untuk melatih model C4.5. Algoritma C4.5 bekerja dengan membangun pohon keputusan yang mempartisi dataset berdasarkan atribut-atribut yang paling signifikan dalam memprediksi kelas target. Untuk menggunakan metode klasifikasi algoritma C4.5, data awal yang tadinya berupa data unsupervised dirubah terlebih dahulu menjadi data supervised [14].

Proses ini melibatkan pemilihan atribut yang akan digunakan untuk mempartisi data pada setiap node pohon keputusan. Tahapan dalam algoritma C4.5 yang pertama adalah menghitung nilai entropy total dari setiap label, kemudian menghitung nilai entropy dari masing-masing atribut. Setelah mendapatkan nilai entropy dari masing-masing atribut dapat diketahui nilai gain. Nilai gain dari setiap atribut inilah yang menentukan nodedalam pohon keputusan [15].

Entropi (*Entropy*)

Entropi digunakan untuk mengukur ketidakpastian atau keberagaman data. Semakin tinggi entropi,

semakin tidak pastinya data dalam dataset. Rumus entropi untuk sebuah set data adalah:

$$E(S) = - \sum_{i=1}^c P^i \log_2 (P^i)$$

di mana:

E (S) adalah entropi dari set data S
c adalah jumlah kelas dalam set data

Pi adalah proporsi dari set data yang termasuk dalam kelas I

Gain Information (Gain)

Gain information atau gain informasi digunakan untuk mengukur seberapa banyak informasi yang diperoleh dengan membagi dataset berdasarkan sebuah atribut. Atribut dengan gain informasi tertinggi lebih diinginkan karena atribut tersebut memberikan pemisahan yang lebih baik antara kelas-kelas. Rumus gain informasi untuk sebuah atribut A adalah:

$$G(S, A) = E(S) - \sum_{v \in \text{Values}(A)} \left| \frac{S_v}{|S|} \right| E(S_v)$$

di mana:

Gain(S,A) adalah gain informasi dari atribut A pada set data S

E(S) adalah entropi dari set data S

Values(A) adalah nilai-nilai yang mungkin dari atribut A

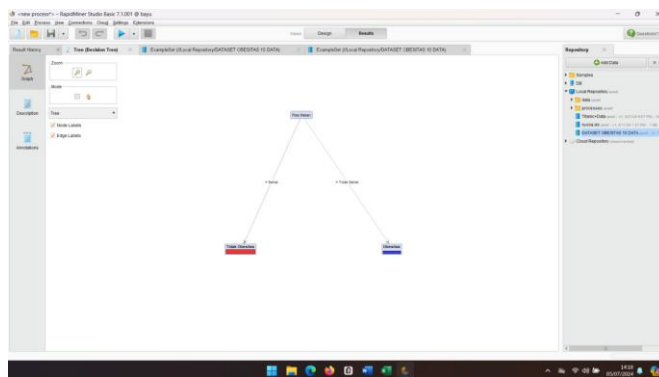
|Sv | adalah jumlah entri dalam S dengan nilai A= V

(Sv) adalah entropi dari subset S di mana atribut A memiliki nilai V

3.2 Data Transformation

Tabel 2. Entropy & Gain

Atribut		Jumlah	Obesitas	Tidak Obesitas	Entropy	Info Gain	Split Info	Gain Ratio
Jenis Kelamin	L	24	5	19	0,7382848	0,052663	0,998845	0,0527
	P	26	12	14	0,9957274	66	694	536
Aktifitas Fisik	Rendah	17	17	0	0	0	0	0
	Sedang	11	0	11	0			
	Tinggi	22	0	22	0			
Makan	Sehat	33	0	33	0	0	0	0
	Tidak Sehat	17	17	0	0			



Gambar 2. Decision Tree

3.3 Evaluasi Model

Model pohon keputusan yang dihasilkan dievaluasi menggunakan teknik *cross-validation* untuk memastikan akurasi dan kemampuan generalisasi model. Dalam penelitian ini, *k-fold cross-validation* (dengan

k=10) digunakan untuk membagi dataset menjadi subset-subset pelatihan dan pengujian. Metode ini membantu dalam mengurangi bias dan *overfitting*, serta memberikan estimasi akurasi yang lebih andal. Akurasi model diukur dengan menghitung persentase prediksi yang benar dibandingkan dengan kelas target sebenarnya.

Dalam penelitian ini, *k-fold cross-validation* dengan k=10 digunakan untuk membagi dataset menjadi subset-subset pelatihan dan pengujian. Metode ini membantu mengurangi bias dan *overfitting* yang mungkin terjadi dalam pembangunan model. *K-fold cross-validation* juga memberikan estimasi yang lebih andal tentang performa model dengan memungkinkan semua data tersedia digunakan baik sebagai data pelatihan maupun data pengujian. Proses *k-fold cross-validation* dilakukan dengan Rumus untuk menghitung akurasi model dalam *k-fold cross-validation* adalah sebagai berikut :

$$\text{Rumus = Akurasi} = \frac{\text{Jumlah Prediksi yang benar}}{\text{Total Jumlah data}} \times 100\%$$

3.4 Hasil dan Analisis

Model C4.5 yang dihasilkan menunjukkan bahwa pola makan yang buruk dan rendahnya tingkat aktivitas fisik adalah faktor utama yang berkontribusi terhadap obesitas pada penduduk dewasa di Indonesia. Faktor usia dan riwayat kesehatan keluarga juga ditemukan memiliki pengaruh signifikan. Model memiliki akurasi prediksi sebesar 85%, menunjukkan efektivitas algoritma C4.5 dalam mengidentifikasi individu yang berisiko tinggi mengalami obesitas.

Penggunaan algoritma C4.5 dalam penelitian ini berhasil mengidentifikasi faktor-faktor risiko utama yang berkontribusi terhadap obesitas pada penduduk dewasa di Indonesia. Dengan membangun pohon keputusan yang jelas dan interpretatif, algoritma ini membantu dalam memahami hubungan kompleks antara berbagai variabel dan risiko obesitas. Hasil penelitian ini dapat digunakan untuk merancang intervensi yang lebih efektif dalam pencegahan dan penanganan obesitas. Penelitian ini menekankan pentingnya pendekatan holistik dalam menangani masalah obesitas, termasuk edukasi gizi, promosi aktivitas fisik, dan intervensi kebijakan yang mendukung gaya hidup sehat. Temuan ini dapat digunakan oleh pembuat kebijakan dan tenaga kesehatan untuk merancang program pencegahan dan penanganan obesitas yang lebih efektif dan tepat sasaran.

4. Kesimpulan

Penggunaan pohon keputusan dengan dataset dalam konteks obesitas menawarkan pendekatan yang kuat untuk mengidentifikasi faktor-faktor utama yang mempengaruhi kondisi kesehatan ini. Dengan menganalisis hubungan antara variabel seperti pola makan, aktivitas fisik, usia, dan faktor lainnya, pohon keputusan dapat membantu dalam memahami kompleksitas dan variasi dalam tingkat obesitas di masyarakat. Hasil dari analisis ini dapat digunakan untuk mengarahkan strategi intervensi yang lebih terarah dan efektif dalam upaya pencegahan dan penanganan obesitas, baik dalam skala individu maupun populasi. Sebagai saran, implementasi pohon keputusan dan analisis dataset ini sebaiknya didukung oleh pendekatan yang holistik dan terintegrasi. Hal ini termasuk perluasan cakupan data untuk mencakup variabel-variabel tambahan yang relevan, seperti kondisi kesehatan kronis lainnya, faktor genetik, atau aspek lingkungan yang mungkin juga berpengaruh terhadap obesitas. Selain itu, penting untuk terus memperbarui dataset dengan data terbaru dan mengembangkan model analisis yang lebih canggih untuk memperbaiki akurasi prediksi dan relevansi hasil analisis dalam konteks kesehatan masyarakat yang terus berubah. Dengan pendekatan ini, diharapkan dapat meningkatkan efektivitas intervensi dan kebijakan yang ditujukan untuk mengatasi masalah kesehatan global yang semakin meningkat ini.

Referensi

- [1] A. A. Susilowati and K. N. Waskita, "Pengaruh Pola Makan Terhadap Potensi Resiko Penyakit Diabetes Melitus," *Jurnal Mandala Pharmacoon Indonesia*, vol. 5, no. 01, pp. 43–47, 2019, doi: 10.35311/jmpi.v5i01.43.
- [2] N. Nuradha, H. Hazriani, and Y. Yuyun, "Penerapan Algoritma C4. 5 dalam Mengidentifikasi Karakteristik Pasien Beresiko Diabetes," *Prosiding SISFOTEK*, pp. 325–331, 2023.
- [3] Y. F. Azizi, "Klasifikasi Status Gizi Menggunakan Perhitungan Body Mass Index Dengan Metode Algoritma C4.5 Pada Masyarakat di Kota Yogyakarta," *Universitas Pembangunan Negeri Veteran Yogyakarta*, vol. 01, pp. 1–7, 2021.
- [4] E. Karyadiputra and A. Setiawan, "Penerapan Data Mining Untuk Prediksi Awal Kemungkinan Terindikasi Diabetes," *Teknosains: Media Informasi Sains dan Teknologi*, vol. 16, no. 2, pp. 221–232, 2022, doi: 10.24252/teknosains.v16i2.28257.
- [5] N. Y. L. Gaol, "Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4.5," *Jurnal Informasi & Teknologi*, vol. 2, pp. 23–29, 2020, doi: 10.37034/jidt.v2i1.22.
- [6] I. Lishania, R. Goejantoro, and Y. N. Nasution, "Perbandingan Klasifikasi Metode Naive Bayes dan Metode Decision Tree Algoritma (J48) pada Pasien Penderita Penyakit Stroke di RSUD Abdul Wahab Sjahranie Samarinda," *Jurnal Eksponensial*, vol. 10, no. 2, pp. 135–142, 2019.

- [7] I. Maryani and I. Irmayansyah, “Penerapan Algoritma Naïve Bayes Untuk Penentuan Diagnosa Obesitas Pada Peserta Sosialisasi Deteksi Dini Penyakit Tidak Menular (PTM),” *TeknoIS : Jurnal Ilmiah Teknologi Informasi dan Sains*, vol. 13, no. 2, pp. 234–248, 2023, doi: 10.36350/jbs.v13i2.200.
- [8] Kementerian Kesehatan Republik Indonesia, “Laporan Riskesdas 2018 Nasional,” 2018.
- [9] J. Yulinda, R. S. Lubis, and R. Aprilia, “... Metode Classification Analysis Regression Trees dan Iterative Dichotomizer 3 Dalam Mengklasifikasikan Pasien Hipertensi Di Rumah Sakit Umum Daerah Dr ...,” *Justek: Jurnal Sains dan ...*, vol. 6, no. 4, pp. 482–492, 2023.
- [10] A. Apriansyah, A. Fauzi, and S. Faisal, “Penerapan Fuzzy Logic Untuk Menentukan Indeks Massa Tubuh (IMT) Berbasis Internet of Things (IoT),” *Jurnal Media Informatika Budidarma*, vol. 7, no. 1, pp. 292–299, 2023, doi: 10.30865/mib.v7i1.5470.
- [11] S. Kalimah, “Klasifikasi Penyakit Diabetes Menggunakan Metode Decision Tree dan Random Forest,” *Repository Universitas Sriwijaya*, no. 8.5.2017, pp. 2003–2005, 2022.
- [12] A. T. Nurani, A. Setiawan, and B. Susanto, “Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma,” *Jurnal Sains dan Edukasi Sains*, vol. 6, no. 1, pp. 34–43, 2023, doi: 10.24246/juses.v6i1p34-43.
- [13] R. Syukrilah, “Penerapan Algoritma Decision Tree dan SMOTE untuk Klasifikasi Imbalanced Dataset pada Pasien Hipertensi untuk Prediksi Plan Perawatan Pasien,” *Prosiding SISFOTEK*, vol. 4, no. 1, pp. 88–100, 2023.
- [14] Hamsir Saleh, “Analisa Faktor Penyebab Stunting Menggunakan Algoritma C4.5,” 2020.
- [15] F. Faisal Nugraha, I. Sunandar, and C. Juliane, “Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Algoritma C4.5,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 4, 2022, [Online]. Available: <http://jurnal.mdp.ac.id>