



Analisis Faktor Risiko Obesitas Pada Individu Menggunakan Algoritma Random Forest Dengan Rapidminer

Naqibuzzahidin^{1*}, Zaehol Fatah²

^{1,2} Teknologi Informasi, Universitas Ibrahimy

^{1*}naqibuzzahidin22@gmail.com, ²zaeholfatah@gmail.com

Abstrak

Obesitas adalah masalah kesehatan global yang ditandai dengan menumpuknya lemak tubuh berlebih, meningkatkan risiko penyakit kronis seperti diabetes dan penyakit jantung. Data Mining sangat efektif untuk mengidentifikasi faktor risiko obesitas pada individu menggunakan algoritma Random Forest yang diimplementasikan dengan RapidMiner, dipilih karena kemampuannya menangani banyak variabel dan mengatasi multikolinearitas. Dataset ini terdiri dari 2.087 entri dengan 15 fitur terkait gaya hidup, pola makan, dan riwayat kesehatan yang diperoleh dari situs Kaggle, dan dibagi menjadi 90% data latih serta 10% data uji untuk evaluasi kinerja model. Hasilnya menunjukkan model mampu mencapai akurasi 79,81%, presisi 78,26%, dan recall 67,86%, dengan faktor dominan yang mempengaruhi risiko obesitas adalah konsumsi makanan tinggi kalori, rendahnya aktivitas fisik, dan riwayat keluarga obesitas. Penelitian ini memberikan wawasan penting bagi upaya pencegahan obesitas melalui pengembangan kebijakan kesehatan dan program edukasi yang lebih efektif, fokus pada perubahan pola hidup untuk mengurangi prevalensi obesitas dan komplikasi

Kata Kunci: Obesitas, Faktor Risiko, Random Forest, RapidMiner, Data Mining.

PENDAHULUAN

Obesitas adalah kondisi medis yang ditandai oleh akumulasi lemak tubuh yang berlebihan, sehingga menyebabkan berat badan seseorang melampaui batas yang dianggap sehat. Kondisi ini biasanya terjadi ketika asupan energi yang diperoleh dari makanan dan minuman melebihi energi yang digunakan atau dibakar oleh tubuh dalam aktivitas sehari-hari. Ketidakseimbangan ini menyebabkan tubuh menyimpan energi berlebih dalam bentuk lemak. Salah satu cara yang umum digunakan untuk mengukur tingkat obesitas adalah melalui Indeks Massa Tubuh (IMT). IMT menghitung rasio antara berat badan dan tinggi badan seseorang, yang kemudian digunakan untuk mengkategorikan status berat badan. Menurut standar kesehatan yang berlaku secara global, seseorang dikatakan memiliki berat badan berlebih jika nilai IMT berada di atas 25,0, dan seseorang dianggap mengalami obesitas ketika nilai IMT melebihi 30,0. Kondisi obesitas tidak hanya memengaruhi penampilan fisik, tetapi juga memiliki dampak serius terhadap kesehatan, meningkatkan risiko berbagai penyakit kronis seperti diabetes tipe 2, penyakit jantung, hipertensi, dan gangguan metabolismik lainnya. Oleh karena itu, obesitas memerlukan penanganan yang serius serta upaya pencegahan yang efektif (Annurullah et al. 2021).

Faktor-faktor yang berkontribusi terhadap obesitas sangat beragam, dan hal ini membuatnya menjadi fenomena yang kompleks. Menurut Teori Klasik HL Blum, terdapat empat faktor utama yang mempengaruhi tingkat kesehatan seseorang, yaitu: 1) perilaku, 2) lingkungan, 3) layanan kesehatan, dan 4) faktor genetik. Keempat faktor ini berinteraksi secara kompleks dan berkontribusi pada kondisi kesehatan individu secara keseluruhan, termasuk risiko obesitas. Perilaku dan gaya hidup, seperti pola makan yang tidak sehat dan kurangnya aktivitas fisik, merupakan faktor utama yang mempengaruhi kemungkinan seseorang mengalami obesitas. Di sisi lain, faktor lingkungan, baik sosial maupun fisik, juga memainkan peran yang penting. Misalnya, ketersediaan akses terhadap makanan sehat dan fasilitas olahraga di lingkungan sekitar seseorang dapat memengaruhi kebiasaan kesehatan individu tersebut. Selain itu, layanan kesehatan yang tidak memadai serta faktor genetik yang diwariskan juga dapat meningkatkan kerentanan seseorang terhadap obesitas (Saraswati et al. 2021).

Untuk memahami lebih dalam faktor-faktor yang berkontribusi terhadap obesitas, penggunaan teknologi dalam analisis data menjadi sangat relevan. Salah satu pendekatan yang menjanjikan adalah penerapan algoritma pembelajaran mesin (machine learning), seperti Random Forest. Algoritma Random Forest didasarkan pada teknik pohon keputusan yang memungkinkan analisis data yang lebih mendalam. Algoritma ini mampu mengidentifikasi pola-pola yang kompleks dan memprediksi hasil berdasarkan berbagai variabel yang terkait. Salah satu keunggulan utama dari Random Forest adalah kemampuannya dalam menangani dataset yang besar dan bervariasi, termasuk data yang memiliki banyak variabel independen. Selain itu, Random Forest juga dapat mengatasi masalah multikolinearitas, yakni ketika variabel independen memiliki korelasi satu sama lain, yang sering kali menjadi kendala dalam analisis data kesehatan. Dengan keunggulan-keunggulan ini, Random Forest menjadi alat yang efektif dalam meneliti faktor-faktor risiko yang berkontribusi terhadap obesitas dan penyakit terkait (Suci Amaliah, Nusrang, and Aswi 2022).

Dalam konteks implementasi algoritma Random Forest, RapidMiner muncul sebagai perangkat lunak analisis data yang dapat mempermudah proses tersebut. RapidMiner adalah perangkat lunak yang dirancang dengan antarmuka visual yang intuitif, memungkinkan pengguna untuk mengimplementasikan algoritma pembelajaran mesin tanpa perlu memiliki kemampuan pemrograman yang mendalam. Dengan menggunakan RapidMiner, analisis data menjadi lebih mudah karena perangkat lunak ini tidak hanya memfasilitasi proses pemrosesan dan visualisasi data, tetapi juga memungkinkan pengguna untuk melakukan eksperimen dengan berbagai teknik pembelajaran mesin secara efisien. Selain itu, RapidMiner juga mendukung integrasi dengan berbagai sumber data yang berbeda, sehingga memudahkan pengguna dalam mengolah data dari berbagai format (Rafi Nahjan, Nono Heryana, and Apriade Voutama 2023).

Penelitian ini bertujuan untuk menganalisis faktor risiko obesitas pada individu dengan menggunakan algoritma Random Forest yang diimplementasikan dalam perangkat lunak RapidMiner. Melalui analisis ini, diharapkan dapat diperoleh wawasan yang lebih mendalam mengenai faktor-faktor utama yang berkontribusi terhadap obesitas, baik dari segi perilaku, lingkungan, maupun faktor genetik. Hasil dari penelitian ini diharapkan dapat membantu dalam penyusunan strategi pencegahan yang lebih tepat dan efektif, yang tidak hanya dapat diterapkan pada level individu, tetapi juga pada level komunitas dan populasi yang lebih luas. Strategi pencegahan tersebut diharapkan mampu menurunkan risiko obesitas secara signifikan dan pada akhirnya mengurangi prevalensi penyakit kronis yang berkaitan dengan kondisi ini.

METODE

Jenis Penelitian

Penelitian ini menerapkan pendekatan deskriptif kuantitatif, yang fokus pada bidang rekayasa perangkat lunak serta pemanfaatan Library Online Research sebagai sumber data utama. Proses pengumpulan data dilakukan melalui studi literatur yang mendalam. Pemilihan metode deskriptif kuantitatif bertujuan untuk memberikan gambaran yang jelas, mendetail, dan obyektif mengenai faktor-faktor risiko yang berkontribusi terhadap obesitas. Metode ini memungkinkan analisis data berdasarkan angka secara sistematis, sehingga peneliti dapat menarik kesimpulan yang tepat berdasarkan informasi numerik yang diperoleh. Selain itu, pendekatan ini mendukung pengelolaan dan penyimpanan data secara terstruktur untuk mengidentifikasi secara lebih akurat faktor-faktor risiko yang berkaitan dengan obesitas (Zulfikar et al. 2022).

Metode Pengumpulan Data

Teknik pengumpulan data adalah metode yang digunakan untuk memperoleh data yang diperlukan dalam penelitian, baik dari subjek maupun sampel. Proses ini sangat penting karena menjadi dasar dalam menyusun instrumen penelitian yang akan digunakan untuk mengumpulkan data terkait faktor risiko obesitas. Pengumpulan data yang akurat dan sesuai prosedur akan menghasilkan informasi yang kredibel, yang penting untuk analisis menggunakan algoritma random forest dalam mengidentifikasi faktor risiko obesitas. Kesalahan dalam metode ini dapat mengurangi keandalan hasil analisis (Ardiansyah, Risnita, and Jailani 2023). Untuk mendapatkan data yang relevan dengan faktor risiko obesitas, peneliti menggunakan metode studi literatur dengan mengandalkan data sekunder yang diperoleh dari situs Kaggle dan data ini digunakan sebagai dasar untuk analisis menggunakan algoritma random forest.

Dataset yang digunakan dalam analisis faktor risiko obesitas pada individu dapat diakses melalui tautan berikut: <https://www.kaggle.com/datasets/lkjotsingh221/obesity-risk-prediction-cleaned>. Dataset ini berisi 2.087 entri dalam format CSV, yang mencakup 15 atribut atau fitur yang berkaitan dengan prediksi kondisi obesitas pada pasien. Data tersebut mengandung informasi diagnostik yang diperlukan untuk mengidentifikasi apakah seseorang menderita obesitas atau tidak. Berikut adalah penjelasan mengenai atribut-atribut dalam dataset, yang disajikan dalam gambar di bawah ini.



A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	Height	Weight	family_history_with_overweight	SCC	MTRANS_Walking	FAVC_z	FCVC_minmax	NCP_z	CAEC_minmax	CH2O_minmax	FAF_minmax	TUE_z	CALC_z	Age_bin_minmax
2	162	640	1	0	0	2,76688E+15	5	4,04704E+16	3,33333E+15	5	0.0	5,50985E+15	1,43903E+16	25
3	152	560	1	1	0	2,76688E+15	10	4,04704E+16	3,33333E+15	10	1.0	1,09272E+16	5,16552E+15	25
4	18	770	1	0	0	2,76688E+15	5	4,04704E+16	3,33333E+15	5	0.6666666666666666	5,50985E+15	2,47214E+15	5
5	18	870	0	0	1	2,76688E+15	10	4,04704E+16	3,33333E+15	5	0.6666666666666666	1,09272E+16	2,47214E+15	75
6	178	898	0	0	0	2,76688E+15	5	2,16412E+16	3,33333E+15	5	0.0	1,09272E+16	5,16552E+15	5
7	162	530	0	0	0	3,61418E+16	5	4,04704E+16	3,33333E+15	5	0.0	1,09272E+16	5,16552E+15	10
8	15	550	1	0	0	3,61418E+16	10	4,04704E+16	3,33333E+15	5	0.3333333333333333	1,09272E+16	5,16552E+15	5
9	164	530	0	0	0	2,76688E+15	5	4,04704E+16	3,33333E+15	5	1.0	1,09272E+16	5,16552E+15	5
10	178	640	1	0	0	3,61418E+16	10	4,04704E+16	3,33333E+15	5	0.3333333333333333	5,50985E+15	2,47214E+15	75
11	172	680	1	0	0	3,61418E+16	5	4,04704E+16	3,33333E+15	5	0.3333333333333333	5,50985E+15	1,43903E+16	5
12	185	1050	1	0	0	3,61418E+16	10	4,04704E+16	6,66667E+15	10	0.6666666666666666	2,19469E+16	5,16552E+15	75
13	172	800	1	1	0	3,61418E+16	5	4,04704E+16	6,66667E+15	5	0.6666666666666666	5,50985E+15	1,43903E+16	25
14	165	560	0	0	0	2,76688E+15	10	4,04704E+16	3,33333E+15	10	0.6666666666666666	1,09272E+16	5,16552E+15	5
15	18	990	0	0	0	3,61418E+16	5	4,04704E+16	3,33333E+15	5	0.6666666666666666	5,50985E+15	2,47214E+15	10
16	177	600	1	0	0	3,61418E+16	10	2,16412E+16	3,33333E+15	0	0.3333333333333333	5,50985E+15	5,16552E+15	5
17	17	660	1	1	0	2,76688E+15	10	4,04704E+16	3,33333E+15	10	5,6666666666666666	5,50985E+15	5,16552E+15	5
18	193	1020	1	0	0	3,61418E+16	5	2,16412E+16	3,33333E+15	0	0.3333333333333333	1,09272E+16	5,16552E+15	75
19	153	780	0	0	0	3,61418E+16	5	2,16412E+16	3,33333E+15	5	0.0	1,09272E+16	1,43903E+16	10
20	171	820	1	0	0	3,61418E+16	10	1,68911E+16	6,66667E+15	0	0.0	1,09272E+16	1,43903E+16	10
21	165	700	1	0	0	2,76688E+15	5	2,16412E+16	3,33333E+15	5	0.0	1,09272E+16	5,16552E+15	5
22	165	800	1	0	1	2,76688E+15	5	4,04704E+16	3,33333E+15	5	1.0	2,19469E+16	1,43903E+16	5
23	165	600	1	0	0	3,61418E+16	10	4,04704E+16	3,33333E+15	5	0.3333333333333333	1,09272E+16	5,16552E+15	5
24	16	820	1	0	0	3,61418E+16	0	2,16412E+16	3,33333E+15	5	0.0	2,19469E+16	5,16552E+15	5
25	185	680	1	0	0	3,61418E+16	5	4,04704E+16	3,33333E+15	5	0.0	5,50985E+15	5,16552E+15	25
26	16	500	1	0	0	2,76688E+15	5	1,68911E+16	6,66667E+15	5	1.0	2,19469E+16	1,43903E+16	25
27	17	650	1	0	1	3,61418E+16	5	2,16412E+16	6,66667E+15	5	0.3333333333333333	2,19469E+16	4,42772E+16	25
28	16	520	0	0	0	3,61418E+16	5	1,68911E+16	6,66667E+15	5	0.6666666666666666	5,50985E+15	5,16552E+15	5
29	175	760	1	1	0	3,61418E+16	10	4,04704E+16	3,33333E+15	5	1.0	5,50985E+15	5,16552E+15	0
30	168	700	0	0	1	3,61418E+16	5	4,04704E+16	3,33333E+15	5	0.6666666666666666	2,19469E+16	2,47214E+15	5
31	177	830	0	0	0	3,61418E+16	0	1,68911E+16	6,66667E+15	10	0.0	5,50985E+15	1,43903E+16	10
32	158	680	1	0	0	2,76688E+15	5	2,16412E+16	3,33333E+15	0	0.3333333333333333	1,09272E+16	5,16552E+15	10
33	177	760	0	0	1	2,76688E+15	5	4,04704E+16	3,33333E+15	10	0.3333333333333333	5,50985E+15	5,16552E+15	75
34	179	900	0	0	0	2,76688E+15	5	2,16412E+16	3,33333E+15	5	0.0	1,09272E+16	5,16552E+15	10
35	165	620	0	0	0	3,61418E+16	5	6,66667E+15	5,0985E+15	5	0.6666666666666666	1,09272E+16	5,16552E+15	5
36	15	650	1	0	0	2,76688E+15	5	4,04704E+16	3,33333E+15	5	0.6666666666666666	2,19469E+16	5,16552E+15	25
37	156	490	0	1	1	3,61418E+16	5	4,04704E+16	3,33333E+15	0	0.6666666666666666	1,09272E+16	1,43903E+16	5
38	16	480	0	0	0	3,61418E+16	5	4,04704E+16	3,33333E+15	0	0.3333333333333333	1,09272E+16	5,16552E+15	25
39	165	670	1	0	0	3,61418E+16	5	4,04704E+16	3,33333E+15	5	0.3333333333333333	5,50985E+15	5,16552E+15	5

Gambar 1. Dataset

Keterangan:

- Height: Tinggi badan (meter).
- Weight: Berat badan (kg).
- family_history_with_overweight: Riwayat keluarga obesitas (1: ya, 0: tidak).
- SCC: Konsumsi makanan berkali tinggi (1: ya, 0: tidak).
- MTRANS_Walking: Transportasi dengan berjalan kaki (1: ya, 0: tidak).
- FAVC_z: Konsumsi makanan tinggi kalori (skor).
- FCVC_minmax: Frekuensi makan sayur.
- NCP_z: Jumlah makanan utama per hari.
- CAEC_minmax: Frekuensi makan camilan.
- CH2O_minmax: Konsumsi air harian.
- FAF_minmax: Frekuensi aktivitas fisik.
- TUE_z: Waktu di depan layar (TV/komputer).
- CALC_z: Konsumsi alkohol.
- Age_bin_minmax: Usia (dikelompokkan).
- NObeyesdad: Kategori obesitas (target).

Data Mining

Data mining adalah proses untuk menemukan korelasi, pola, dan tren yang signifikan dengan menganalisis data dalam jumlah besar yang tersimpan di repositori, menggunakan teknik pengenalan pola serta metode statistik dan matematika. Data mining bertujuan untuk mengungkap pola dan informasi menarik dari data yang berukuran besar. Istilah data mining sering kali dihubungkan dengan istilah lain seperti *knowledge discovery* atau *pattern recognition*. Istilah *knowledge discovery* atau penemuan pengetahuan dianggap tepat karena tujuan utama dari data mining adalah untuk mengungkap pengetahuan tersembunyi di dalam data. Sementara itu, istilah *pattern recognition* atau pengenalan pola cocok digunakan karena proses ini berfokus pada penemuan pola yang tersimpan dalam kumpulan data besar (Teknologi et al. 2021).

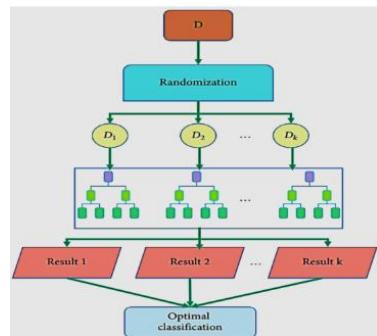
Klasifikasi

Klasifikasi adalah proses menilai data untuk menempatkannya ke dalam salah satu kelas yang tersedia. Dalam proses ini, model dibuat berdasarkan data pelatihan yang kemudian digunakan untuk mengklasifikasikan data baru. Klasifikasi dapat diartikan sebagai proses pembelajaran atau pelatihan terhadap fungsi target yang memetakan setiap kumpulan atribut (fitur) ke salah satu kelas yang ada. Tujuan dari sistem klasifikasi adalah mengklasifikasikan semua data secara akurat, namun kinerjanya tidak selalu 100% sempurna. Oleh karena itu, diperlukan pengukuran untuk mengevaluasi performa sistem klasifikasi, yang biasanya dilakukan dengan menggunakan matriks konfusi (Arrohman and Fatah 2024).

Algoritma Random Forest

Random Forest adalah algoritma *Supervised Learning* yang digunakan untuk klasifikasi dan regresi. Algoritma ini bekerja dengan metode pengantongan, bukan dengan meningkatkan model. Pohon-pohon dalam Random Forest

beroperasi secara paralel tanpa adanya interaksi saat proses pembentukan pohon berlangsung (Jollyta et al. 2023). Gambar dibawah ini memberikan ilustrasi mengenai algoritma Random Forest sesuai definisi tersebut.



Gambar 2. Skema Algoritma Random Forest

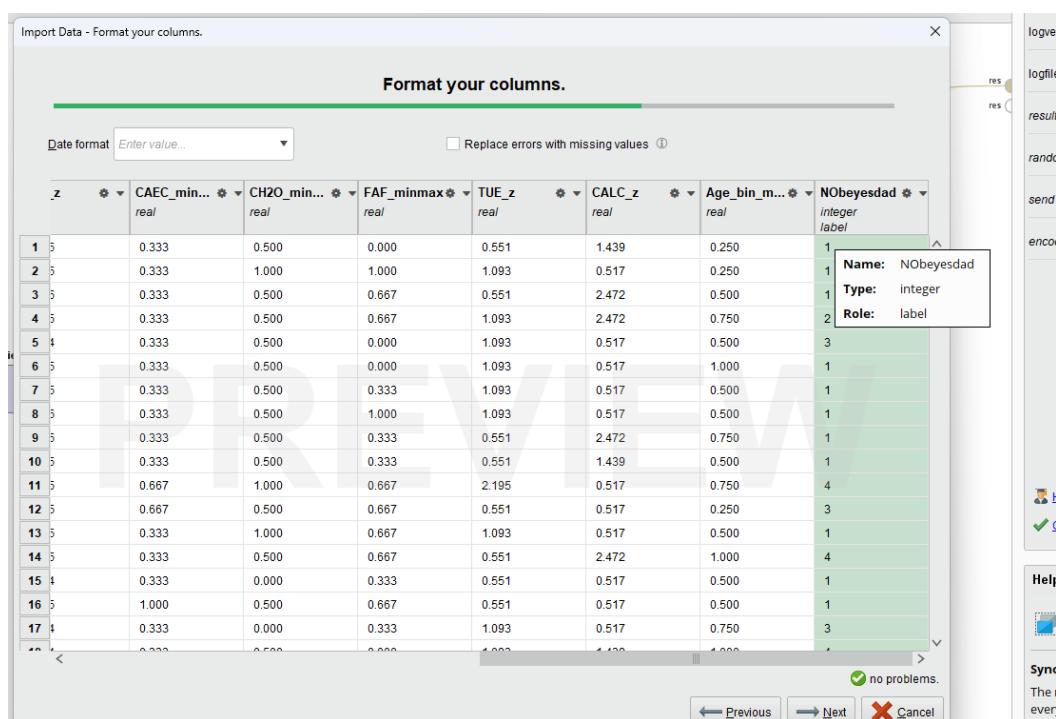
Faktor Risiko Obesitas

Penyebab utama obesitas adalah penumpukan lemak berlebih dalam tubuh yang disebabkan oleh peningkatan asupan energi yang tidak seimbang dengan kebutuhan energi. Ketidakseimbangan ini dipengaruhi oleh berbagai faktor, termasuk biologis, sosial, budaya, politik, lokal, dan ekonomi global. Kerangka biopsikososial untuk memahami obesitas menunjukkan bahwa faktor biologis, psikososial, dan perilaku berperan signifikan dalam menentukan status berat badan anak. Beberapa faktor yang berkontribusi meliputi karakteristik remaja seperti jenis kelamin, usia, berat saat lahir, pola makan, perilaku aktivitas fisik, gaya pengasuhan, dan karakteristik keluarga, serta gaya hidup dan lingkungan rekreasi. Faktor-faktor ini saling berinteraksi dan mempengaruhi pola makan, jenis makanan yang dikonsumsi, dan tingkat aktivitas fisik individu. Meskipun pilihan ini bersifat individu, banyak dari keputusan tersebut dipengaruhi oleh keluarga, terutama pada anak-anak. Hubungan antara keluarga dan orang tua dapat memengaruhi kecenderungan individu untuk mengalami kelebihan berat badan atau obesitas melalui berbagai mekanisme, termasuk faktor genetik, budaya, psikologis, perilaku, dan sosial, atau kombinasi dari faktor-faktor tersebut (Pratama 2023).

HASIL DAN PEMBAHASAN

Penelitian ini dilakukan menggunakan salah satu model klasifikasi, yaitu Random Forest. Model ini diterapkan menggunakan perangkat lunak RapidMiner versi 10.3 untuk mempermudah proses analisis data. Dataset yang digunakan dalam penelitian ini adalah "*Estimation of obesity levels based on eating habits and physical condition*", yang diperoleh dari situs Kaggle. Data ini berisi 15 atribut yang relevan untuk mengklasifikasikan individu berdasarkan risiko obesitas, dengan kategori individu yang mengalami obesitas atau tidak, yang dapat digunakan sebagai parameter utama. Proses menggunakan visualisasi di RapidMiner dilakukan melalui beberapa langkah berikut:

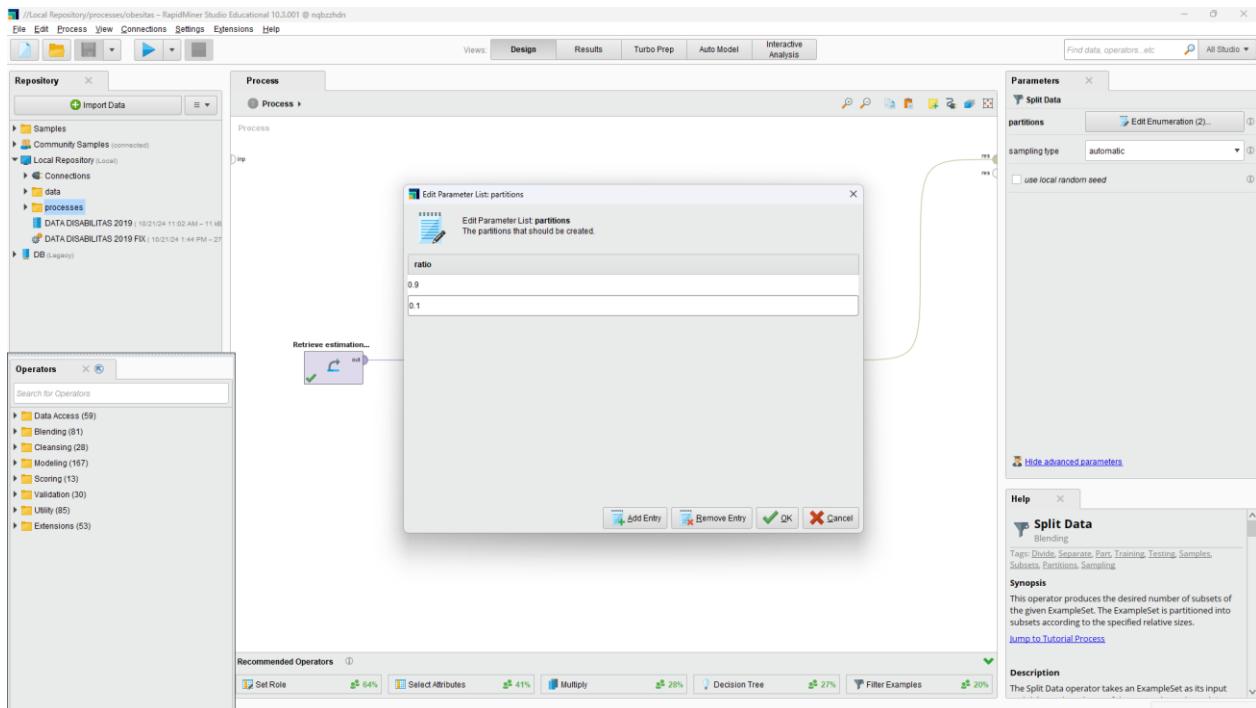
Model Data Mining



Gambar 3. Operator Pemanggil Data

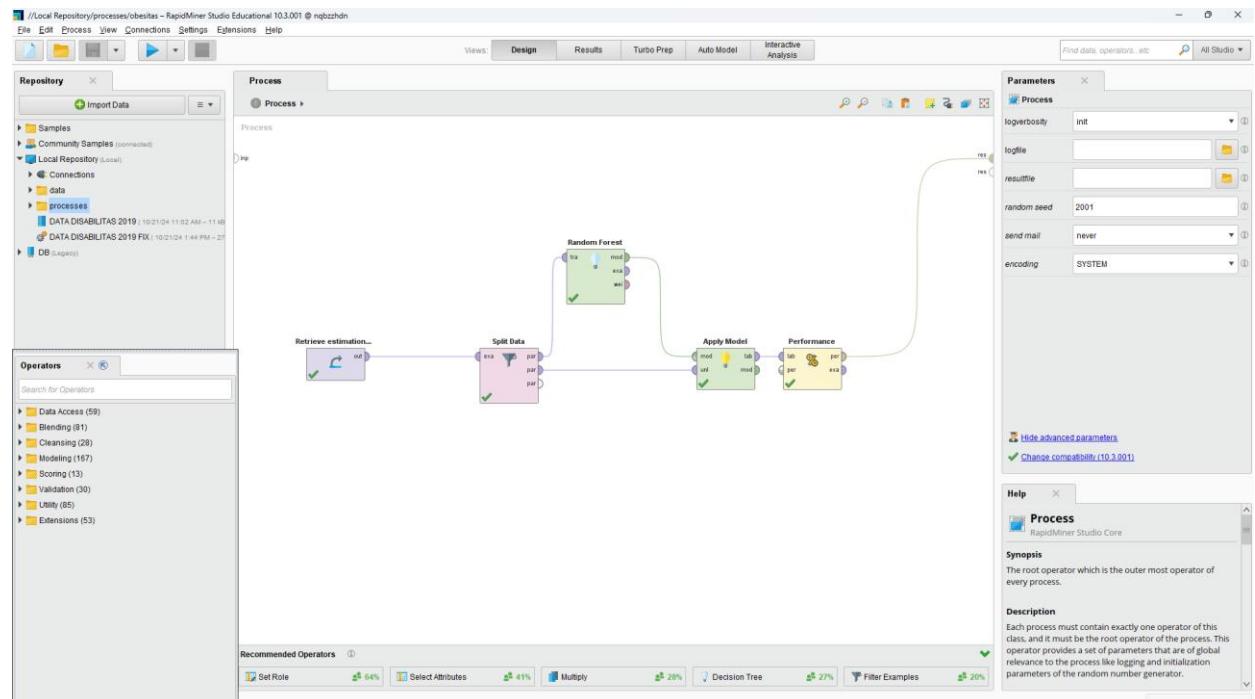


Operator Read CSV digunakan untuk memuat dataset yang disimpan dalam format CSV. Pada bagian ini, kita juga menentukan atribut yang akan digunakan sebagai label dalam dataset yang dipakai.



Gambar 4. Split Data

Split data berfungsi untuk memisahkan dataset menjadi data latih (training data) dan data uji (testing data) (Oktafiani, Hermawan, and Avianto 2023). Langkah ini sangat penting untuk menghindari overfitting dan mengevaluasi kemampuan model dalam mengidentifikasi faktor risiko obesitas pada individu. Operator Split Data dapat digunakan untuk membagi dataset, misalnya 90% untuk data latih dan 10% untuk data uji, guna memastikan bahwa model Random Forest dapat menggeneralisasi dengan baik terhadap data baru, bukan hanya data yang digunakan untuk pelatihan.



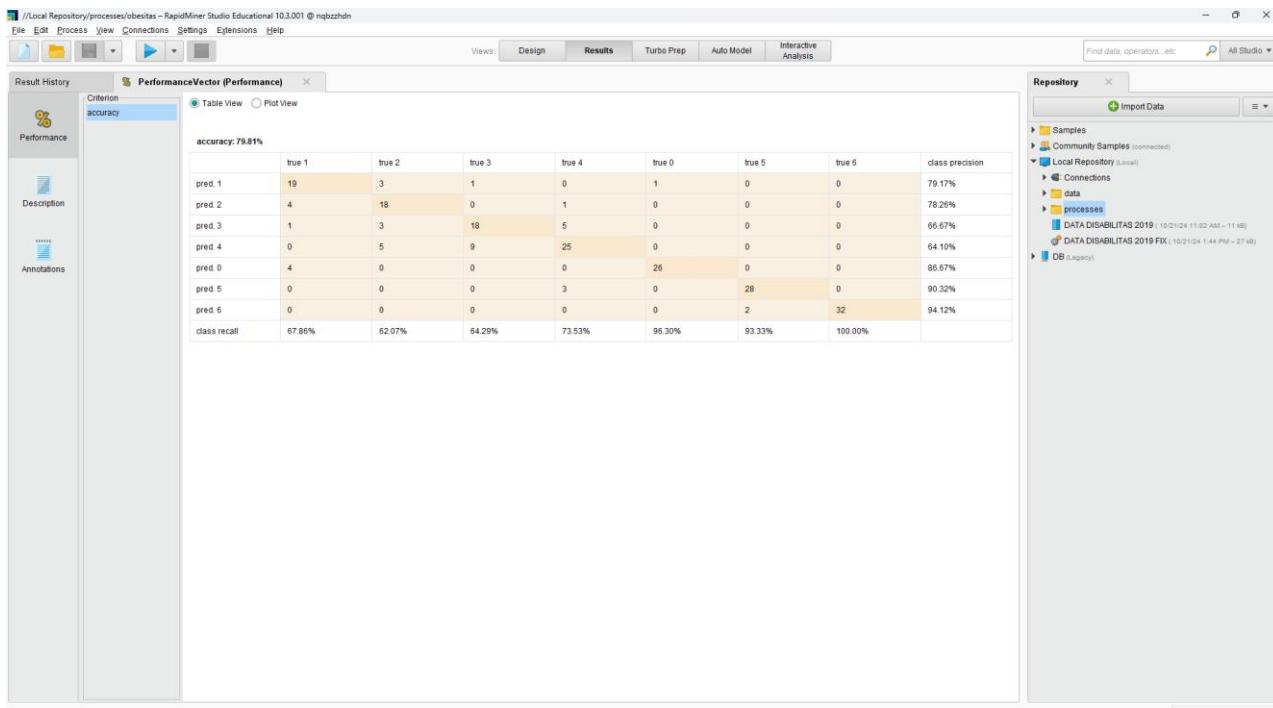
Gambar 5. Model Random Forest

Implementasi data mining dengan algoritma Random Forest menggunakan aplikasi RapidMiner versi 10.3 dilakukan melalui beberapa langkah seperti yang ditunjukkan pada gambar diatas. Tahap selanjutnya adalah menghubungkan semua operator yang telah disiapkan.

Operator Apply Model merupakan komponen penting yang digunakan untuk menerapkan model yang telah dilatih ke dataset baru, memungkinkan prediksi atau klasifikasi dilakukan pada data yang belum dikenal berdasarkan model yang dibangun dari data latih. Setelah melatih model Random Forest menggunakan data latih, operator Apply Model dapat digunakan untuk menerapkan model tersebut pada data uji guna memprediksi risiko obesitas pada individu.

Operator Performance digunakan untuk mengevaluasi kinerja model yang telah dilatih dan diterapkan. Ini adalah alat penting untuk menilai efektivitas model dalam memprediksi atau mengklasifikasikan faktor risiko obesitas, dengan menyediakan berbagai metrik seperti akurasi, presisi, dan recall untuk menilai kualitas prediksi yang dihasilkan oleh model setelah diterapkan pada data uji (Al-Giffary and Martanto 2024).

Hasil Akurasi



Gambar 6. Hasil Akurasi

Berdasarkan gambar diatas, Penerapan algoritma Random Forest dalam menganalisis faktor risiko obesitas menghasilkan nilai akurasi sebesar 79,81%. Dari matriks kebingungan yang ditampilkan, kita dapat melihat distribusi prediksi dan nilai aktual untuk berbagai kelas. Misalnya, pada kelas "true 1," terdapat 19 prediksi yang benar dan 3 yang salah sebagai kelas lain. Hasil evaluasi juga menunjukkan nilai presisi dan recall untuk setiap kelas. Presisi tertinggi dicapai pada kelas "true 6" dengan 100% recall, sementara presisi terendah ditemukan pada kelas "true 3" sebesar 66,67%. Evaluasi kinerja menggunakan berbagai metrik ini memberikan gambaran tentang seberapa baik model Random Forest dalam memprediksi faktor risiko obesitas di antara individu yang dianalisis. Dengan mempertimbangkan hasil ini, model ini menunjukkan kemampuan yang baik dalam mengklasifikasikan data, meskipun ada beberapa kelas yang memerlukan peningkatan dalam hal presisi dan recall untuk mendapatkan hasil yang lebih akurat dalam analisis risiko obesitas.

KESIMPULAN

Penelitian ini menemukan bahwa konsumsi makanan tinggi kalori, kurangnya aktivitas fisik, dan riwayat keluarga obesitas adalah faktor dominan yang meningkatkan risiko obesitas. Pola makan tidak seimbang dan gaya hidup sedentari juga berkontribusi secara signifikan. Algoritma Random Forest terbukti efektif dalam mengidentifikasi faktor risiko utama obesitas, memberikan wawasan penting bagi kebijakan kesehatan dan program pencegahan obesitas yang lebih tepat sasaran, seperti peningkatan aktivitas fisik dan pola makan sehat. Meski hasil ini kuat, penelitian dibatasi oleh ukuran dataset yang kecil; studi lanjutan dengan data yang lebih besar dan beragam direkomendasikan untuk hasil yang lebih umum.

UCAPAN TERIMA KASIH

Dengan penuh rasa syukur, saya mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, baik secara langsung maupun tidak langsung, dalam penyelesaian penelitian ini. Ucapan terima kasih saya tujuhan kepada pembimbing dan dosen yang telah memberikan bimbingan, arahan, dan pengetahuan berharga selama proses penelitian berlangsung.

Tidak lupa, saya juga berterima kasih kepada keluarga dan teman-teman yang selalu memberikan dukungan moral serta semangat tanpa henti. Bantuan dan dorongan dari Anda semua telah menjadi motivasi besar bagi saya untuk menyelesaikan penelitian ini dengan sebaik-baiknya.

Semoga hasil penelitian ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan, khususnya di bidang kesehatan dan data mining. Saya juga berharap bahwa penelitian ini dapat menjadi langkah awal untuk penelitian lebih lanjut di masa depan.

DAFTAR PUSTAKA

- Al-Giffary, Farhan Rizky, and Martanto Martanto. 2024. "Klasifikasi Kelulusan Siswa Tahun 2024 Menggunakan Metode Decision Tree (Studi Kasus Sma Islam Alazhar 5 Cirebon)." *Jurnal Manajemen Informatika Jayakarta* 4(2): 195.
- Annurullah, Gifta Alifa, Maulyda Shakeela Jasmine, Nandita Ardrafitri Saraswati, and Yabsutur Rizka. 2021. "Faktor Risiko Obesitas Pada Pekerja Kantoran: A Systematic Review." *Jurnal Kesehatan Tambusai* 2(2): 80–88.
- Ardiansyah, Rismita, and M. Syahran Jailani. 2023. "Teknik Pengumpulan Data Dan Instrumen Penelitian Ilmiah Pendidikan Pada Pendekatan Kualitatif Dan Kuantitatif." *Jurnal IHSAN : Jurnal Pendidikan Islam* 1(2): 1–9.
- Arrohman, Supri, and Zaehol Fatah. 2024. "Gudang Jurnal Multidisiplin Ilmu Prediksi Diabetes Menggunakan Algoritma Klasifikasi K-Nearest Neighbors (K-NN) Pada Perempuan Indian Pima." 2: 220–26.
- Jollyta, D, A Hajjah, E Haerani, and M Siddik. 2023. *Algoritma Klasifikasi Untuk Pemula Solusi Python Dan RapidMiner*. Deepublish. <https://books.google.co.id/books?id=y84TEQAAQBAJ>.
- Oktafiani, Rian, Arief Hermawan, and Donny Avianto. 2023. "Pengaruh Komposisi Split Data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning." *Jurnal Sains dan Informatika* 9(April): 19–28.
- Pratama, Bangkit Ary. 2023. "Literature Review: Faktor Risiko Obesitas Pada Remaja Di Indonesia." *Indonesian Journal on Medical Science* 10(2).
- Rafi Nahjan, Muhammad, Nono Heryana, and Apriade Voutama. 2023. "Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell." *JATI (Jurnal Mahasiswa Teknik Informatika)* 7(1): 101–4.
- Saraswati, Siwi Kurnia et al. 2021. "Literature Review : Faktor Risiko Penyebab Obesitas." *Media Kesehatan Masyarakat Indonesia* 20(1): 70–74.
- Suci Amaliah, Muhammad Nusrang, and Aswi Aswi. 2022. "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konijiwa Bantaeng." *VARIANSI: Journal of Statistics and Its application on Teaching and Research* 4(3): 121–27.
- Teknologi, Jurnal et al. 2021. "ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19." 2(2): 100–108.
- Zulfikar, Z, E S Podungge, M I Saleh, and ... 2022. "Penerapan Data Mining Untuk Memprediksi Tingkat Kelulusan Siswa Menggunakan Algoritma Neural Network." *Jurnal Elektronik Sistem* ... 5(1): 7–13. <http://jesik.web.id/index.php/jesik/article/view/91>.