



Prediksi Risiko Obesitas Remaja Mengacu pada Konsumsi dan Olahraga Dengan Random Forest

Inkka Kavita¹, Marselo Charly², Ajai Shan³, Dirga Arefa Wibowo⁴,

^{1,2,3,4} Universitas Satya Terra Bhinneka

Korespondensi penulis: inkkavita06@email.com

Abstract. *Teenager's obesity is public health issue that needs attention because of the increasing risk of other diseases such as diabetes. This study purposes to build a system that can predict the risk of teenager's obesity. The data is in the form of a secondary dataset that has 17 features that include eating habits, physical activity, and others. Random Forest is used because of its ability to handle high-dimensional data and produce accurate classifications. This system is console-based, where users can add their lifestyle data and get results in the form of obesity predictions with low, medium, and high levels. The results of the model evaluation show very good results, which 90% accurate and high consistency, recall, and f1-score values. This method shows stable and competitive performance compared to other algorithms such as Decision Tree and KNN. So the results are expected to be used as a learning and prevention tool to detect the risk of obesity in adolescents early on.*

KEYWORDS: *Obesity, Lifestyle, Risk, Prediction, Random Forest*

Abstrak. Obesitas remaja adalah isu kesehatan di masyarakat yang perlu diperhatikan karena bertambahnya risiko penyakit lain seperti diabetes. Penelitian ini ditujukan agar dapat memprediksi risiko obesitas remaja. Data berbentuk dataset sekunder yang memiliki 17 fitur yang mencakup kebiasaan makan, aktivitas fisik, dan lainnya. Random Forest digunakan karena kelebihannya dalam menangani data berdimensi besar dan menghasilkan klasifikasi yang akurat. Sistem ini berbasis console, di mana user dapat menambahkan data gaya hidup mereka dan memperoleh hasil berupa prediksi obesitas dalam tiga tingkatan: rendah, sedang, dan tinggi. Hasil evaluasi kinerja model memaparkan hasil yang sangat baik, dengan keakuratan 90%-an dan nilai konsistensi, nilai recall, dan nilai f1-score yang cukup tinggi. Metode ini menunjukkan kinerja stabil dan kompetitif dibandingkan dengan algoritma lain seperti Decision Tree dan KNN. Sehingga hasil penelitian ini diharapkan dapat dimanfaatkan sebagai alat belajar dan pencegahan untuk mendeteksi risiko obesitas pada remaja sejak dini.

KATA KUNCI: Obesitas, Gaya Hidup, Risiko, Prediksi, Random Forest

1. LATAR BELAKANG

Obesitas menjadi isu kesehatan masyarakat yang semakin mengkhawatirkan, terutama bagi kaum remaja. Banyak masalah kesehatan yang berisiko karena obesitas terutama akumulasi lemak *visceral* untuk perkembangan hormon insulin. Terutama disebabkan oleh pola konsumsi dan kurangnya olahraga yang dilakukan. Tetapi penyebabnya tidak terbatas pada itu saja, tetapi juga terkait dengan gangguan hormonal akibat *disfungsi tiroid* seperti *hipotiroidisme*. Saat ini, remaja lebih sering menghabiskan banyak waktu untuk *screen time* dan mengonsumsi makanan tinggi kalori dan gula. Obesitas adalah masalah kesehatan yang banyak ditemui di masyarakat Indonesia. Menurut Siloam Hospitals (2010), data Risdas tahun 2016 menunjukkan bahwa angka obesitas dewasa di Indonesia sebesar 20,7%. Angka prevalensi terlihat meningkat dari tahun 2013 sebesar 15,4%. Maka kemajuan teknologi yaitu *random forest* dapat dilakukan. “Random forest adalah algoritma pembelajaran yang diusulkan oleh Breiman [Mach. Learn. 45 (2001) 5–32] yang menggabungkan beberapa pohon keputusan

acak dan menggabungkan prediksi mereka dengan merata-ratakan”, (Erwan Scornet, Gérard Biau, Jean-Philippe Vert (2015)). Karena pendekatan yang sama dapat diadaptasi untuk memprediksi risiko obesitas pada remaja berdasarkan pola makan dan olahraga secara fisik, dengan memperhatikan faktor-faktor lain seperti pola asuh dan lingkungan. Informasi kesehatan dan pilihan gaya hidup remaja dapat diproses oleh teknologi ini untuk memperkirakan risiko obesitas secara akurat dan mudah dipahami.

2. KAJIAN TEORITIS

Obesitas adalah kondisi yang ditandai dengan tertimbunnya lemak yang dapat mengganggu kesehatan. Ketidaksamaan energi yang masuk dan yang dibutuhkan tubuh untuk melakukan tugas biologis sehingga menyebabkan obesitas. Prediktor obesitas dibuat dengan metode random forest untuk menyeimbangkan dan mencegah obesitas. Secara umum, metode ini dianggap akurat dan dapat menangani struktur data yang kompleks, ruang fitur yang besar, dan ukuran sampel yang kecil. “Karena random forest beroperasi dengan membentuk beberapa pohon keputusan selama pelatihan dan menghasilkan klasifikasi atau regresi dari masing-masing pohon”, (GeeksforGeeks(2024, February 15)). Sebagai acuan, digunakan penelitian "Implementasi Algoritma K-Nearest Neighbor (KNN), Random Forest, Naïve Bayes, dan Decision Tree untuk Mengklasifikasikan Tingkat Obesitas". Penelitian ini bermaksud untuk melihat perbandingan empat metode klasifikasi untuk menentukan tingkat obesitas yaitu: *K-Nearest Neighbor (KNN), Decision Tree, Naïve Bayes, dan Random Forest*.

3. METODE PENELITIAN

Penelitian ini bersifat kuantitatif yang memakai konsep pendekatan *data mining* melalui implementasi sistem klasifikasi berbasis machine learning dengan memakai dataset sekunder yang bersumber dari Kaggle. Dimana kami menggunakan data-based approach dan mengikuti konsep supervised learning berdasarkan konsep kecerdasan buatan. Menggunakan data sekunder yang bersumber dari penelitian sebelumnya. Data sekunder telah disiapkan untuk analisis dengan langkah-langkah pembersihan dan pre-processing data yang meliputi normalisasi, pengisian data hilang, dan encoding variable. Tahapannya terdiri dari:

- Persiapan Data,
- Pembersihan Data,
- Analisis Data, dan
- Pembangunan Model.

Dimana terdapat langkah lanjutan untuk analisis data, antara lain:

- Preprocessing,
- Split Dataset,
- Modeling,
- Evaluasi Model, dan
- Penyimpanan Model.

4. HASIL DAN PEMBAHASAN

Dataset yang berisi 17 atribut yang terdiri dari berbagai fitur yang mencerminkan aspek gaya hidup dan kondisi personal seperti gender, umur, tinggi badan, berat badan, pola makan, aktivitas fisik dan konsumsi air.

Pengujian dilakukan berdasarkan 423 data uji yang telah melalui proses encoding dan normalisasi.

Hasil ini sejalan dengan Rifaldi et al. (2022), yang melaporkan akurasi Random Forest sebesar 94,82% dalam klasifikasi obesitas berdasarkan variabel gaya hidup. Hal ini akan didukung dengan evaluasi kinerja model yang menggunakan metrik klasifikasi yaitu precision, recall, dan f1-score.

Tabel 1. Hasil ini sejalan dengan Rifaldi

Data Uji	Precision	Recall	F1 Score	Support
0	1.00	0.93	0.96	54
1	0.83	0.98	0.90	58
2	0.96	0.96	0.96	70
3	1.00	0.98	0.99	60
4	1.00	0.98	0.99	65
5	0.96	0.90	0.93	58
6	0.96	0.95	0.96	58

Berdasarkan hasil evaluasi berikut, terlihat bahwa model memiliki kinerja yang tinggi pada hampir seluruh kelas, dengan nilai f1-score rata-rata sebesar **0.96**, baik pada perhitungan rata-rata makro maupun rata-rata tertimbang. Performa yang tinggi ini memaparkan bahwa model klasifikasi yang digunakan mempunyai kapabilitas yang andal dalam mendeteksi tingkat obesitas berdasarkan fitur-fitur yang dipakai dalam pelatihan. Dan ini merupakan perbandingan empat model yaitu XGBoost, Random Forest, Decision Tree, dan KNN.

Tabel 2. Implikasi random forest pada penelitian

Metode	Akurasi	Precision	Recall	F1 Score
XGBoost	0.96	0.96	0.96	0.96
Random Forest	0.95	0.95	0.95	0.95
Decision Tree	0.94	0.94	0.94	0.94
KNN	0.82	0.82	0.82	0.81

Implikasi random forest pada penelitian ini yaitu:

- Secara teoritis dapat memperkuat efektivitas Random Forest dalam klasifikasi data medis multivariat.
- Secara terapan dapat digunakan sekolah atau layanan kesehatan remaja untuk deteksi dini dan edukasi pola hidup sehat.

Analisis Dataset

Dataset yang digunakan yaitu 2111 baris data dan 17 atribut, yang merupakan 4 atribut utama dan 13 variabel. Target atau label dari dataset adalah kategori obesitas, yang dibagi menjadi beberapa kelas:

- normal_weight,
- overweight_level_i,
- obesity_type_ii,
- obesity_type_i,
- insufficient_weight.

Data telah melalui proses preprocessing, yang meliputi:

- Encoding variabel kategorikal dengan *LabelEncoder*.
- Normalisasi data numerik.
- Pembagian dataset.
- Penanganan data hilang jika ada.

1. Deskripsi Atribut Dataset

Terdapat 4 kolom spesifik yang terdiri dari **Gender**, **Age**, **Height**, **Weight**. **Gender** adalah pilihan jenis kelamin yaitu female dan male. **Age** adalah spesifikasi umur yang menampung nilai bulat positif. **Height** adalah spesifikasi tinggi badan yang ditampung dalam bentuk meter. **Weight** adalah spesifikasi berat badan yang ditampung dalam kilogram.

a) Atribut Lifestyle

Variabel yang digunakan adalah sebagai berikut:

- Gender adalah pilihan jenis kelamin.
- Usia adalah pilihan usia.
- Berat_badan adalah pilihan berat badan dalam kg.
- Tinggi_badan adalah pilihan tinggi badan dalam cm.
- Family_history adalah riwayat keturunan secara genetik.
- FAVC adalah keseringan makan makanan tinggi kalori.
- FCVC adalah banyak sayuran yang masuk ke dalam makanan berat.
- NCP adalah berapa kali biasanya kamu makan dalam sehari.
- CAEC adalah apakah kamu konsumsi makanan selain makanan berat.
- SMOKE adalah apakah kamu punya kebiasaan merokok.
- CH2O berapa banyak air yang kamu minum dalam sehari.
- SCC adalah perhatian kamu akan jumlah kalori yang kamu konsumsi.
- FAF adalah seberapa sering kamu melakukan aktivitas fisik.
- TUE adalah waktu yang dihabiskan untuk gadget.
- CALC adalah seberapa sering minum alkohol.
- MTRANS adalah jenis transportasi yang biasanya kamu digunakan.
- Obesity_level adalah target hasil keseluruhan apakah tergolong obesitas atau tidak.

Evaluasi Model

Penilaian lanjutan dilakukan dengan metrik seperti akurasi, precision, recall, dan f1-score. Lalu lakukan perbandingan random forest dengan jenis model lainnya. Hasil yang diperoleh yaitu:

- Akurasi: 96%
- Rata-rata makro F1-score: 0.96
- Rata-rata berat F1-score: 0.96

Hasil evaluasi menunjukkan bahwa model mempunyai performa yang tinggi pada hampir seluruh kelas, dengan nilai f1-score rata-rata sebesar **0.96**, baik pada perhitungan rata-rata makro maupun rata-rata tertimbang. Nilai **akurasi keseluruhan** juga mencapai **96%**, yang menunjukkan bahwa model mampu mengelompokkan data dengan sangat baik. Secara khusus, *Overweight Level II* menunjukkan kinerja tertinggi dengan *precision*, *recall*, dan *f1-score* sebesar 1.00. Hal ini memaparkan bahwa model mampu mendeteksi seluruh data pada kelas ini secara akurat tanpa kesalahan. Di sisi lain, *Overweight Level I* memiliki nilai f1-score

terendah sebesar 0.88, meskipun nilai tersebut masih tergolong baik dalam konteks klasifikasi multi-kelas. Menjelaskan hasil evaluasi model (akurasi, precision, recall), grafik confusion matrix, dll.

Hasil Random Forest

Random Forest dipilih sebagai algoritma utama untuk melakukan klasifikasi tingkat obesitas berdasarkan sejumlah variabel input seperti kebiasaan makan, aktivitas fisik, dan data antropometri. Berdasarkan hasil evaluasi yang telah disajikan sebelumnya, Random Forest menunjukkan kinerja yang sangat bagus dengan nilai akurasi sebesar 95.27%, precision sebesar 95.21%, recall sebesar 95.20%, dan f1-score sebesar 95.20%. Jika dibandingkan dengan algoritma lain yang digunakan sebagai pembanding, yaitu *XGBoost*, *Pohon Keputusan*, dan *K-Nearest Neighbors (KNN)*, *Random Forest* memberikan hasil yang kompetitif dan stabil. *XGBoost* memang sedikit lebih unggul nilai akurasinya, namun selisihnya tidak terlalu signifikan jika dibandingkan dengan *Random Forest*. Selain itu, *Random Forest* cenderung lebih mudah untuk diinterpretasikan dan tidak memerlukan tuning parameter yang kompleks seperti *XGBoost*.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil membangun model klasifikasi status obesitas menggunakan *Random Forest Classifier* dengan akurasi tinggi (96%). Model dapat mengenali berbagai kategori obesitas berdasarkan kombinasi fitur lifestyle dan karakteristik personal. Namun, untuk pengembangan selanjutnya disarankan menggunakan teknik validasi silang untuk menghindari overfitting. Penambahan data dari sumber lain dan eksplorasi model lain seperti *XGBoost* atau *LightGBM* juga dapat dilakukan untuk meningkatkan kinerja. Visualisasi lebih lanjut terhadap pengaruh setiap fitur terhadap output dapat digunakan untuk memberikan insight bagi intervensi kesehatan.

DAFTAR PUSTAKA

- (n.d.). *Introduction of Random Forest Classifier to ZigBee Device Network Security*. Retrieved June 14, 2025, JSTOR database.
- Alpiansah, A. B., & Ramdhani, Y. (2023). Optimasi fitur dengan forward selection pada estimasi tingkat obesitas menggunakan Random Forest. *Sistemasi: Jurnal Sistem Informasi*, 12(3), 860–873.
- Badan Kebijakan Kesehatan Kementerian Kesehatan RI. (n.d.). Laporan hasil survei.

- GeeksforGeeks. (2024, February 15). *What are the advantages and disadvantages of Random Forest?*. GeeksforGeeks. June 14, 2025
- Global Observatory for Physical Activity. (n.d.). Country cards.
- Hafid, W., & Hanafi, S. (2019). Hubungan aktivitas fisik dan konsumsi fast food dengan kejadian obesitas pada remaja. *Kampurui Jurnal Kesehatan Masyarakat*, 1(1), 6–10.
- Journal of Medical Systems. (n.d.). *Journal of Medical Systems*. June 14, 2025
- Jurnal Akuntansi dan Keuangan. (n.d.). *Jurnal Akuntansi dan Keuangan*. 14 Juni 2025.
- Kurdanti, W., Suryani, I., Syamsiatun, N. H., Siwi, L. P., Adityanti, M. M., Mustikaningsih, D., & Sholihah, K. I. (2015). Faktor-faktor yang mempengaruhi kejadian obesitas pada remaja. *Jurnal Gizi Klinik Indonesia*, 11(4), 179–190.
- Muchlinski, D., Siroky, D., He, J., & Kocher, M. (2016). *Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data*. *Political Analysis*, 24(1), 87–103.
- Rifaldi, A., Anang, L., & Satrio, I. D. (2022). Implementasi algoritma k-nearest neighbor (KNN), Random Forest, Naive Bayes dan Decision Tree untuk mengklasifikasikan tingkat obesitas. *RG Journal of Applied Science*, 2(2).
- Salam, A. (2010). Faktor risiko kejadian obesitas pada remaja [Undergraduate thesis, Universitas Hasanuddin]. Siloam Hospitals. (n.d.). Apa itu obesitas? Siloam Hospitals.
- Scornet, E., Biau, G., & Vert, J.-P. (2015). *Consistency of random forests*. *The Annals of Statistics*, 43(4), 1716–1741.
- Scornet, E., Biau, G., & Vert, J.-P. (2015). *Consistency of random forests*. *The Annals of Statistics*, 43(4), 1716–1741.
- Shi, T., & Horvath, S. (2006). *Unsupervised learning with random forest predictors*. *Journal of Computational and Graphical Statistics*, 15(1), 118–138. doi:10.1198/106186006X94072
- UNICEF Indonesia. (2022). Analisis lanskap kelebihan berat badan dan obesitas di Indonesia. United Nations Children's Fund.
- Universitas Sumatera Utara. (2022). Faktor-faktor yang berhubungan dengan obesitas pada remaja di SMA Negeri 5 Medan Tahun 2022 [Undergraduate thesis, Universitas Sumatera Utara]. Repositori USU.
- World Health Organization. (2023). Obesity and overweight.