



Prediksi Risiko Obesitas Remaja Berdasarkan Pola Makan dan Aktivitas Fisik Menggunakan Algoritma Random Forest

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

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

Korespondensi penulis: inkkavita06@gmail.com

Abstract. Obesity in adolescents is a growing public health concern due to its association with various chronic diseases such as type 2 diabetes, hypertension, and metabolic disorders. This study aims to develop an obesity risk prediction system for adolescents based on dietary patterns and physical activity data using the Random Forest algorithm. The data was obtained from a secondary dataset available on the Kaggle platform, comprising 2,111 entries and 17 features covering dietary habits, physical activity, and anthropometric characteristics. The Random Forest method was chosen for its ability to handle high-dimensional data and produce accurate classifications. The developed system is a console-based application where users can input their lifestyle data and receive obesity risk predictions in three levels: low, moderate, and high. Model performance evaluation showed excellent results with an accuracy of 95%, as well as consistently high precision, recall, and F1-score values. Compared to other algorithms such as KNN and Decision Tree, Random Forest demonstrated competitive and stable performance. The results of this study are expected to be utilized as an educational and preventive tool for early detection of obesity risk in adolescents.

Keywords: Adolescent Obesity, Dietary Patterns, Physical Activity, Risk Prediction, Random Forest, Machine Learning, Data Mining

Abstrak. Obesitas pada remaja merupakan masalah kesehatan masyarakat yang semakin mengkhawatirkan karena berisiko menimbulkan berbagai penyakit kronis seperti diabetes tipe 2, hipertensi, hingga gangguan metabolik. Penelitian ini bertujuan untuk membangun sebuah sistem prediksi risiko obesitas remaja berbasis data pola makan dan aktivitas fisik menggunakan algoritma Random Forest. Data diperoleh dari dataset sekunder yang tersedia di platform Kaggle, dengan total 2111 entri dan 17 fitur yang mencakup kebiasaan makan, aktivitas fisik, serta karakteristik antropometri. Metode Random Forest digunakan karena kemampuannya dalam menangani data berdimensi tinggi dan menghasilkan klasifikasi yang akurat. Sistem yang dikembangkan bersifat konsol, di mana pengguna dapat memasukkan data gaya hidup mereka dan memperoleh prediksi risiko obesitas dalam tiga tingkatan: rendah, sedang, dan tinggi. Evaluasi performa model menunjukkan hasil yang sangat baik dengan akurasi mencapai 95%, serta nilai precision, recall, dan f1-score yang konsisten tinggi. Dibandingkan dengan algoritma lain seperti KNN dan Decision Tree, Random Forest menunjukkan performa kompetitif dan stabil. Hasil penelitian ini diharapkan dapat dimanfaatkan sebagai alat bantu edukatif dan preventif untuk deteksi dini risiko obesitas pada remaja.

Kata kunci: Obesitas Remaja, Pola Makan, Aktivitas Fisik, Prediksi Risiko, Random Forest, Machine Learning, Data Mining

1. LATAR BELAKANG

Obesitas menjadi isu kesehatan masyarakat yang semakin serius dan berbahaya, terutama bagi kaum remaja. Banyak masalah kesehatan, seperti diabetes mellitus tipe 2 yang berisiko karena meningkatnya insiden obesitas terutama akumulasi lemak *visceral* yang berperan signifikan dalam perkembangan resistensi insulin. Terutama disebabkan oleh pola makan yang buruk dan kurangnya olahraga.

Tetapi penyebabnya tidak hanya terbatas pada pola makan yang buruk dan kurangnya aktivitas fisik, tetapi juga terkait dengan gangguan hormonal akibat *disfungsi tiroid*, seperti

hipotiroidisme. Saat ini, remaja lebih sering menonton televisi atau menggunakan ponsel, serta mengonsumsi makanan tinggi kalori dan gula.

Oleh karena itu, deteksi dini sangatlah penting. Memanfaatkan kemajuan teknologi yaitu algoritma *random forest* adalah salah satu pendekatan yang dapat diterapkan. Karena pendekatan serupa dapat diadaptasi untuk memprediksi risiko obesitas pada remaja berdasarkan pola makan dan aktivitas fisik, dengan mempertimbangkan faktor-faktor lain seperti pola asuh dan lingkungan sosial.

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 didefinisikan sebagai penyakit kronis kompleks yang ditandai dengan timbunan lemak berlebihan yang dapat mengganggu kesehatan. Faktor-faktor risiko utama meliputi pola makan tidak sehat yang bergantung pada makanan instan, aktivitas fisik rendah, serta faktor genetik dan lingkungan. Diagnosis kelebihan berat badan dan obesitas dilakukan dengan mengukur berat dan tinggi badan seseorang serta menghitung indeks massa tubuh (IMT). Dalam konteks prediksi, metode Random Forest dikenal efektif untuk klasifikasi karena kemampuannya mengatasi data berdimensi tinggi dan kompleks serta bersifat non-parametrik. Beberapa penelitian sebelumnya menunjukkan bahwa Random Forest dapat menghasilkan akurasi tinggi dalam prediksi status obesitas dibandingkan metode lain seperti KNN, Naive Bayes, dan Decision Tree seperti penelitian Yang berjudul “Implementasi Algoritma K-Nearest Neighbor (KNN), Random Forest, Naïve Bayes dan Decision Tree untuk Mengklasifikasikan Tingkat Obesitas”.

3. METODE PENELITIAN

Penelitian ini merupakan studi kuantitatif dengan pendekatan supervised learning. Data diperoleh dari dataset Kaggle yang mencakup atribut seperti usia, jenis kelamin, tinggi,

Tabel 1. dataset Kaggle

Gender	Age	Height	Weight	family_history	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	Obesity
Female	21	1.62	64	yes	no	2	3	Sometimes	no	2	no	0	1	no	Public_Transportation	Normal_Weight
Female	21	1.52	56	yes	no	3	3	Sometimes	yes	3	yes	3	0	Some times	Public_Transportation	Normal_Weight

Male	23	1.8	77	yes	no	2	3	Som e time s	no	2	no	2	1	Freque ntly	Public_ Transporta tion	Normal_ Weight
Male	27	1.8	87	no	no	3	3	Som e time s	no	2	no	2	0	Freque ntly	Walking	Overwei ght_ Level_I
Male	22	1.78	89.8	no	no	2	1	Som e time s	no	2	no	0	0	Some times	Public_ Transporta tion	Overwei ght_ Level_II

berat, kebiasaan makan, dan aktivitas fisik.

Proses analisis meliputi :

-Preprocessing (encoding dan normalisasi),

Tabel 2. encoding dan normalisasi

Langkah	Penjelasan
Menghapus nilai kosong (null)	Tidak ditemukan nilai null di dataset.
Mengonversi kolom kategorikal menjadi huruf kecil	Semua nilai string seperti 'Gender', 'family_history', dsb. dikonversi ke huruf kecil.
Label encoding untuk kolom target	Kolom 'Obesity' diubah menjadi nilai numerik untuk klasifikasi.
Split data menjadi training dan testing	80% data digunakan untuk pelatihan, 20% untuk pengujian.

-Pembagian data (80% pelatihan, 20% pengujian),

-Penerapan Random Forest Classifier, dan

-Evaluasi menggunakan metrik yaitu akurasi, precision, recall, dan F1-score.

Model disimpan menggunakan joblib untuk memungkinkan penggunaan ulang.

4. HASIL DAN PEMBAHASAN

Dataset terdiri dari 2.111 baris dan 17 atribut, dengan label target yaitu kategori obesitas. Model Random Forest menunjukkan performa tinggi dengan akurasi sebesar 95%, macro average F1-score 0.95, dan weighted average F1-score 0.95.

Dibandingkan dengan KNN, Decision Tree, dan XGBoost, Random Forest menampilkan kestabilan hasil dan interpretasi yang lebih sederhana. Kelas overweight level II memiliki performa klasifikasi tertinggi, sementara overweight level I sedikit lebih rendah namun tetap di atas 0.88 untuk f1-score.

Hasil ini menunjukkan efektivitas Random Forest dalam mendeteksi berbagai level obesitas berdasarkan gaya hidup remaja.

	precision	recall	f1-score	support
0	1.00	0.98	0.99	56
1	0.89	0.92	0.90	62
2	0.97	0.95	0.96	78
3	0.97	0.97	0.97	58
4	1.00	1.00	1.00	63
5	0.88	0.88	0.88	56
6	0.92	0.94	0.93	50
accuracy			0.95	423
macro avg	0.95	0.95	0.95	423
weighted avg	0.95	0.95	0.95	423

Gambar 1. efektivitas Random Forest

5. KESIMPULAN DAN SARAN

Metode Random Forest dipilih sebagai algoritma utama untuk melakukan klasifikasi tingkat obesitas berdasarkan sejumlah variabel input seperti kebiasaan makan, aktivitas fisik, dan data antropometri.

Penelitian ini membuktikan bahwa algoritma Random Forest efektif digunakan dalam sistem prediksi risiko obesitas remaja, dengan akurasi hingga 95%. Sistem ini dapat membantu remaja, orang tua, dan pihak sekolah dalam memahami risiko obesitas dan mengambil langkah preventif.

```
Choose Files ObesityDataSet.csv
• ObesityDataSet.csv(text/csv) - 263646 bytes, last modified: 4/5/2025 - 100% done
Saving ObesityDataSet.csv to ObesityDataSet (10).csv
/usr/local/lib/python3.11/dist-packages/xgboost/core.py:158: UserWarning: [03:44:08] WARNING: /workspace/src/learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

warnings.warn(msg, UserWarning)

=== Hasil Evaluasi Model ===
      Accuracy Precision  Recall  F1 Score
XGBoost    0.962175  0.962662  0.962175  0.962060
Random Forest 0.952719  0.953615  0.952719  0.953023
Decision Tree 0.945626  0.946027  0.945626  0.945690
KNN         0.820331  0.815933  0.820331  0.811322
```

Gambar 2. algoritma Random Forest

Namun, jika dibandingkan dengan algoritma lain yang digunakan sebagai pembanding, yaitu XGBoost, Decision Tree, dan K-Nearest Neighbors (KNN), Random Forest memberikan hasil yang kompetitif dan stabil.

XGBoost memang sedikit lebih unggul dengan akurasi tertinggi yaitu 96.21%, namun selisihnya tidak terlalu signifikan jika dibandingkan dengan Random Forest. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi teknik validasi silang serta menguji model lain seperti LightGBM untuk perbandingan performa.

Penambahan data primer juga dapat meningkatkan generalisasi model. Dengan mempertimbangkan performa tinggi dan kestabilannya, Random Forest dapat disimpulkan sebagai algoritma yang layak dan andal untuk tugas klasifikasi obesitas ini.

Jumlah data: (2111, 17)

Contoh data:

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	\
0	Female	21.0	1.62	64.0		yes	no	2.0
1	Female	21.0	1.52	56.0		yes	no	3.0
2	Male	23.0	1.80	77.0		yes	no	2.0
3	Male	27.0	1.80	87.0		no	no	3.0
4	Male	22.0	1.78	89.8		no	no	2.0

	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	\
0	3.0	Sometimes	no	2.0	no	0.0	1.0	no	
1	3.0	Sometimes	yes	3.0	yes	3.0	0.0	Sometimes	
2	3.0	Sometimes	no	2.0	no	2.0	1.0	Frequently	
3	3.0	Sometimes	no	2.0	no	2.0	0.0	Frequently	
4	1.0	Sometimes	no	2.0	no	0.0	0.0	Sometimes	

	MTRANS	NOBeyesdad
0	Public_Transportation	Normal_Weight
1	Public_Transportation	Normal_Weight
2	Public_Transportation	Normal_Weight
3	Walking	Overweight_Level_I
4	Public_Transportation	Overweight_Level_II

Gambar 3. Random Forest

DAFTAR REFERENSI

- (n.d.). *Introduction of Random Forest Classifier to ZigBee Device Network Security*. Retrieved June 14, 2025, JSTOR database. <https://www.jstor.org/stable/26487106>
- 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. <https://doi.org/10.30736/sistemasi.v12i3.860>
- Badan Kebijakan Kesehatan Kementerian Kesehatan RI. (n.d.). Laporan hasil survei. <https://www.badankebijakan.kemkes.go.id/laporan-hasil-survei/>
- GeeksforGeeks. (2024, February 15). *What are the advantages and disadvantages of Random Forest?*. GeeksforGeeks. June 14, 2025 <https://www.geeksforgeeks.org/what-are-the-advantages-and-disadvantages-of-random-forest/>
- Global Observatory for Physical Activity. (n.d.). Country cards. <https://new.globalphysicalactivityobservatory.com/>
- 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 <https://link.springer.com/journal/10916>
- Jurnal Akuntansi dan Keuangan. (n.d.). *Jurnal Akuntansi dan Keuangan*. 14 Juni 2025, dari <https://www.jahonline.org/>

- 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. <https://doi.org/10.22146/ijcn.22894>
- 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. <https://doi.org/10.1093/pan/mpv024>
- 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). <https://doi.org/10.31219/osf.io/rg456>
- Salam, A. (2010). Faktor risiko kejadian obesitas pada remaja [Undergraduate thesis, Universitas Hasanuddin]. Siloam Hospitals. (n.d.). Apa itu obesitas? Siloam Hospitals. <https://www.siloamhospitals.com/informasi-siloam/artikel/apa-itu-obesitas>
- Scornet, E., Biau, G., & Vert, J.-P. (2015). *Consistency of random forests. The Annals of Statistics*, 43(4), 1716–1741. <https://doi.org/10.1214/15-AOS1321>
- Scornet, E., Biau, G., & Vert, J.-P. (2015). *Consistency of random forests. The Annals of Statistics*, 43(4), 1716–1741. <https://www.jstor.org/stable/43556658>
- 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 <https://www.jstor.org/stable/27594168>
- UNICEF Indonesia. (2022). Analisis lanskap kelebihan berat badan dan obesitas di Indonesia. United Nations Children's Fund. <https://www.unicef.org/indonesia/media/15581/file/AnalisisLanskapKelebihanBeratBadanObesitasdiIndonesia.pdf>
- 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. <https://repositori.usu.ac.id/handle/123456789/43496>
- World Health Organization. (2023). Obesity and overweight. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>