

Prediksi Keputusan Memulai Diet Pada Usia Pemuda di Jakarta Pendekatan Machine Learning

Fais Jefli^{1*}

¹Politeknik Statistika STIS, DKI Jakarta, Indonesia

*Correspondance author: 212313072@stis.ac.id

ABSTRAK

Obesitas merupakan permasalahan global yang juga dihadapi oleh banyak negara berkembang, termasuk Indonesia. Tingginya prevalensi obesitas sentral pada kelompok usia muda (16–30 tahun), khususnya di DKI Jakarta yang mencatatkan tingkat obesitas sentral tertinggi secara nasional, menunjukkan adanya paradoks. Secara teoritis, kelompok usia muda di wilayah perkotaan seperti Jakarta seharusnya lebih aktif secara fisik. Sehingga penelitian ini bertujuan untuk memprediksi keputusan pemuda di Jakarta pada tahun 2024 untuk memulai diet dengan menggunakan pendekatan *machine learning*. Tiga algoritma yang dibandingkan dalam penelitian ini adalah Naive Bayes, XGBoost dan regresi logistic sebagai *baseline*. Hasil analisis menunjukkan bahwa model XGBoost memiliki performa terbaik dalam memprediksi keputusan pemuda untuk memulai diet, ditunjukkan oleh nilai AUC tertinggi. Berdasarkan analisis nilai SHAP, empat variabel yang memberikan kontribusi terbesar terhadap prediksi adalah umur, jenis kelamin, komentar negatif, dan pergaulan sehat. Nilai SHAP positif pada variabel pergaulan sehat menunjukkan bahwa lingkungan keluarga atau teman yang menerapkan pola hidup sehat meningkatkan kemungkinan prediksi pemuda tersebut untuk memulai diet dari 205 responden penelitian.

Keywords: Diet; Machine Learning; XGBoost

ARTICLE INFO

Submission received: 24 October 2025

Accepted: 31 August 2026

Revised: 18 August 2026

Published: 31 August 2026

Available on: <https://doi.org/10.32493/sm.v8i2.54171>

StatMat: Jurnal Statistika dan Matematika is licenced under a Creative Commons Attribution-ShareAlike 4.0 International License.

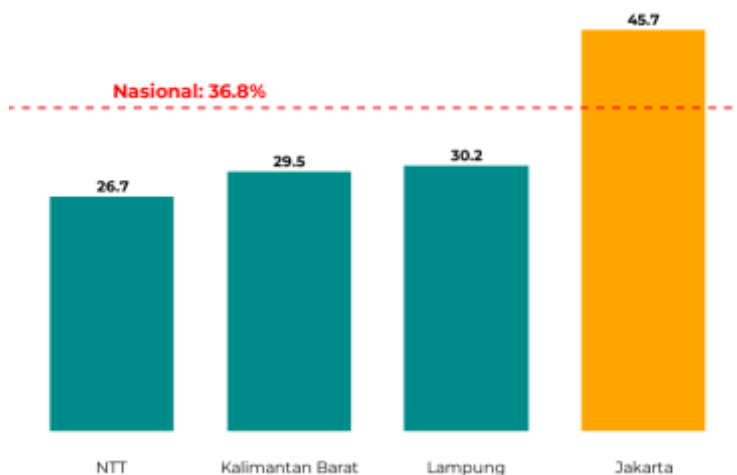
PENDAHULUAN

Obesitas merupakan kondisi penumpukan lemak berlebih akibat ketidakseimbangan antara energi yang dikonsumsi dan energi yang digunakan. Masalah ini menjadi salah satu epidemik global yang masih mengkhawatirkan khususnya di negara berkembang seperti Indonesia (Kemenkes, 2015). Negara-negara berkembang mengalami beban ganda malnutrisi yang masih terus berlangsung, sementara sistem yang ada belum sepenuhnya mampu menangani permasalahan tersebut secara efektif (Kemenkes, 2023a).

Salah satu indikator obesitas adalah obesitas sentral. Menurut WHO (2000), obesitas sentral adalah kondisi tubuh yang memiliki lemak berlebih di perut. Obesitas sentral lebih berkaitan dengan risiko kesehatan dibandingkan indikator obesitas umumnya. Obesitas menyebabkan munculnya berbagai penyakit degeneratif seperti diabetes melitus, hipertensi, dan kardiovaskular (Prastiwi et al., 2019). Selain memberikan dampak terhadap peningkatan penyakit degeneratif, masalah ini juga menyebabkan kerugian ekonomi akibat biaya

pengobatan yang tidak sedikit (Kemenkes, 2023a).

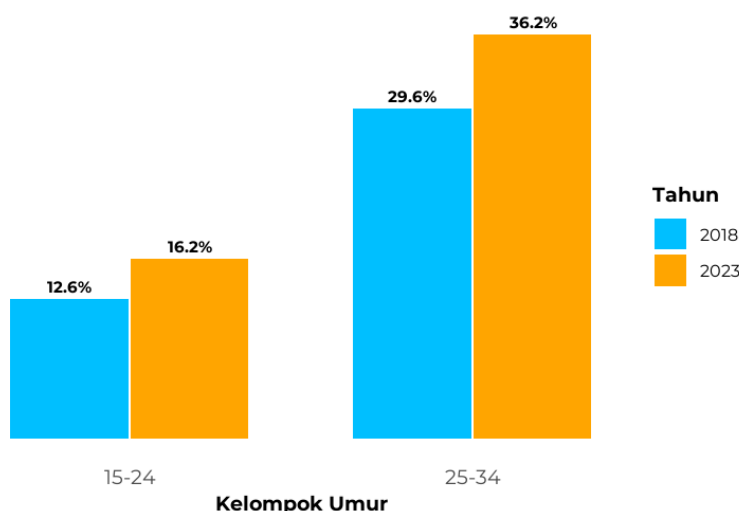
Dalam kurun waktu sepuluh tahun terakhir, terjadi peningkatan obesitas sentral yang cukup signifikan di Indonesia, yaitu dari 31% pada tahun 2018 sampai dengan 36,8% pada tahun 2023. Kondisi ini menunjukkan bahwa masalah obesitas sentral perlu mendapatkan perhatian lebih serius untuk mencegah dampak yang lebih besar di masa mendatang (Kemenkes, 2019 ; Kemenkes, 2023b).



Sumber : Publikasi Kemenkes 2023, diolah

Grafik 1. Prevalensi Obesitas Sentral Pada Penduduk Umur ≥ 15 Tahun Menurut Provinsi Tahun 2023

Berdasarkan Grafik 1, prevalensi obesitas sentral tertinggi tercatat di Provinsi DKI Jakarta, sedangkan tiga provinsi dengan prevalensi terendah ditampilkan sebagai perbandingan. Terlihat bahwa prevalensi obesitas sentral di Jakarta jauh lebih tinggi dibandingkan angka nasional, bahkan hampir dua kali lipat lebih tinggi dibandingkan provinsi dengan prevalensi terendah. Diperkirakan hampir 1 dari 2 penduduk berusia ≥ 15 tahun di Jakarta mengalami obesitas sentral.



Sumber : Publikasi Kemenkes 2018 dan 2023, diolah

Grafik 2. Prevalensi Obesitas Sentral Berdasarkan Kelompok Umur di Indonesia

Berdasarkan Grafik 2, terlihat bahwa dalam kurun waktu sepuluh tahun terjadi

peningkatan prevalensi obesitas sentral pada kelompok usia 15–24 tahun dan 25–34 tahun hingga tahun 2023 (Kemenkes, 2019; Kemenkes, 2023b). Padahal, rentang usia tersebut termasuk dalam usia produktif, khususnya usia 16–30 tahun yang sering dikaitkan dengan fase kepemudaan dan seharusnya menjadi masa ideal untuk melakukan aktivitas fisik. Namun, justru pada kelompok usia ini tetap terjadi peningkatan prevalensi obesitas sentral. Kondisi ini menunjukkan bahwa terdapat berbagai faktor yang menyebabkan kelompok usia tersebut memengaruhi keputusan untuk menjaga kondisi tubuhnya, misalnya melalui diet.

Berdasarkan paparan tersebut, penelitian ini dibuat untuk memprediksi keputusan memulai diet pada usia pemuda dengan berbagai metode *machine learning*, yaitu naive bayes dan XGBoost yang terkenal akan kemampuan prediksinya yang akurat. Peneliti juga membandingkan dengan model regresi logistik. Adapun tujuan dari penelitian ini adalah sebagai berikut

1. Memprediksi keputusan memulai diet pada kelompok usia pemuda di Jakarta dengan model *machine learning* terbaik.
2. Mengetahui arah prediksi dari faktor-faktor tersebut terhadap keputusan memulai diet usia pemuda.

Penelitian ini diharapkan dapat memberikan manfaat secara akademik dengan memperluas kajian penggunaan metode *machine learning* dalam bidang kesehatan, khususnya dalam meningkatkan kemampuan prediksi terhadap suatu kejadian.

METODOLOGI

Ruang Lingkup dan Sumber Data

Penelitian ini memanfaatkan variabel-variabel demografis, sosial, psikologis, dan kesehatan untuk memprediksi keputusan memulai diet pada individu usia pemuda (16–30 tahun) di Provinsi Jakarta. Kategori diet dalam penelitian ini memodifikasi dari definisi Persatuan Ahli Gizi Indonesia (PERSAGI). Unit analisis dalam penelitian ini terdiri dari 205 responden valid yang merupakan individu berusia pemuda dan berdomisili di Provinsi Jakarta. Responden yang telah melakukan diet selama lebih dari tiga bulan tidak termasuk dalam unit analisis karena tidak memenuhi kriteria inklusi, yaitu responden yang baru memulai diet. Seluruh variabel dalam penelitian ini bersumber dari data primer, yaitu data mentah hasil survei yang diperoleh melalui platform *tsurvey.id*.

Operasional Variabel

Variabel respons pada penelitian ini adalah keputusan usia pemuda untuk memulai diet. Penelitian ini memfokuskan analisis pada individu usia pemuda. Responden yang tidak melakukan diet diklasifikasikan ke dalam kategori 0, sedangkan mereka yang melakukan diet dengan durasi maksimal tiga bulan diklasifikasikan ke dalam kategori 1. Responden yang tidak memenuhi kedua kriteria tersebut dikeluarkan dari unit analisis. Adapun penjelasan dari variabel-variabel prediktor yang digunakan dalam penelitian ini adalah sebagai berikut:

Tabel 1. Variabel Prediktor yang Menentukan Keputusan Diet Usia Pemuda

Variabel	Keterangan	Kategori	Rujukan
Aspek Demografis			
Jenis Kelamin	Jenis kelamin	1: Perempuan 0: Laki-laki	Davy et al. (2006)

Umur	Umur	Kontinu	Haynos et al. (2018)
Aspek Sosial			
Pergaulan Sehat	Dalam 3 bulan terakhir, apakah ada keluarga/teman sekitar Anda yang sedang menjalani hidup sehat?	1: Ya 0: Tidak	Chung et al., (2017)
Konten sehat	Dalam 3 bulan terakhir, apakah Anda sering melihat konten di media sosial seputar kesehatan?	1: Ya 0: Tidak	Oktrisia et al. (2021)
Aspek Psikologis			
Kepercayaan Diri	Dalam 3 bulan terakhir, apakah merasa puas dengan tampilan tubuh Anda?	1: Ya 0: Tidak	Oktrisia et al. (2021)
Komentar Negatif	Dalam 3 bulan terakhir, apakah Anda pernah mendapat komentar negatif dari orang lain mengenai tubuh Anda?	1: Ya 0: Tidak	Matera et al. (2013)
Aspek Kesehatan			
Riwayat Penyakit	Apakah Anda memiliki riwayat penyakit terkait pola makan?	1: Ya 0: Tidak	Tan et al. (2022)

Regresi Logistik Biner

Regresi logistik biner adalah model regresi untuk memprediksi kelas variabel respons yang berupa kejadian biner (Ya/Tidak). Untuk menghindari permasalahan hasil prediksi yang dapat melebihi batas 0 dan 1 pada regresi linear konvensional, digunakan fungsi logistik sebagai berikut (James et al., 2021):

$$p(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} \quad (1)$$

Persamaan (1) akan disesuaikan dengan fungsi *log odds* sebagai berikut:

$$\ln\left(\frac{p(x)}{1 - p(x)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (2)$$

Naive Bayes

Naive Bayes classifier adalah metode klasifikasi yang didasarkan pada teorema *bayes*. Model ini disebut sebagai “naive” karena menggunakan asumsi yang sederhana yaitu variabel prediktor dianggap saling bebas secara kondisional terhadap kelas target. Adapun formula dari peluang posterior dalam *naive bayes* sebagai berikut (James et al., 2021):

$$\Pr(Y = k | X = x) = \frac{\pi_k \cdot f_{k1}(x_1) \cdot f_{k2}(x_2) \cdots f_{kp}(x_p)}{\sum_{l=1}^K \pi_l \cdot f_{l1}(x_1) \cdot f_{l2}(x_2) \cdots f_{lp}(x_p)} \quad (3)$$

XGBoost Classifier

XGBoost (Extreme Gradient Boosting) merupakan metode *machine learning* yang dikenal karena kinerjanya yang unggul, skalabilitas, serta fleksibilitasnya. Metode ini menerapkan algoritma *gradient boosting*, yakni pendekatan dengan menggabungkan sejumlah *weak learner* menjadi satu model prediktif yang lebih kuat. Proses pelatihan model dilakukan secara bertahap dengan tujuan meminimalkan fungsi kerugian (loss function), dan parameter model diperbarui melalui teknik *gradient descent* (Ok dan Emmanuel, 2025).

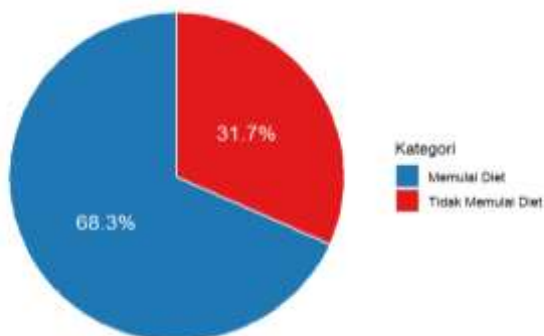
Shapley Values

Shapley value (SHAP) adalah metode yang digunakan untuk melihat kontribusi tiap variabel prediktor terhadap output prediksi model. Nilai SHAP positif menunjukkan bahwa variabel prediktor berkontribusi terhadap prediksi “sukses” begitu juga sebaliknya (Rodríguez-Pérez dan Bajorath, 2020). Adapun tahapan analisis dalam penelitian ini adalah sebagai berikut:

1. Dataset dibagi menjadi dua bagian, yaitu 85% untuk data latih dan 15% untuk data uji.
2. Pada data latih akan dibangun tiga model *machine learning* yaitu regresi logistik, *naïve bayes*, dan XGboost
3. Lalu, akan dihitung AUC untuk ketiga model pada data uji. Model dengan nilai AUC tertinggi adalah model terbaik
4. Model terbaik akan diinterpretasikan dengan SHAP

HASIL DAN PEMBAHASAN

Gambaran Umum



Grafik 3. Persentase Keputusan Memulai Diet 205 Pemuda di Jakarta Tahun 2024.

Berdasarkan Grafik 3, dari 205 pemuda yang berdomisili di Jakarta, sebanyak 68,3% diketahui sedang memulai diet, sementara sisanya belum memulai diet. Hal ini menunjukkan bahwa kesadaran akan pentingnya menjaga pola makan dan kesehatan tubuh mulai tumbuh di kalangan pemuda Jakarta.



Membangun Model

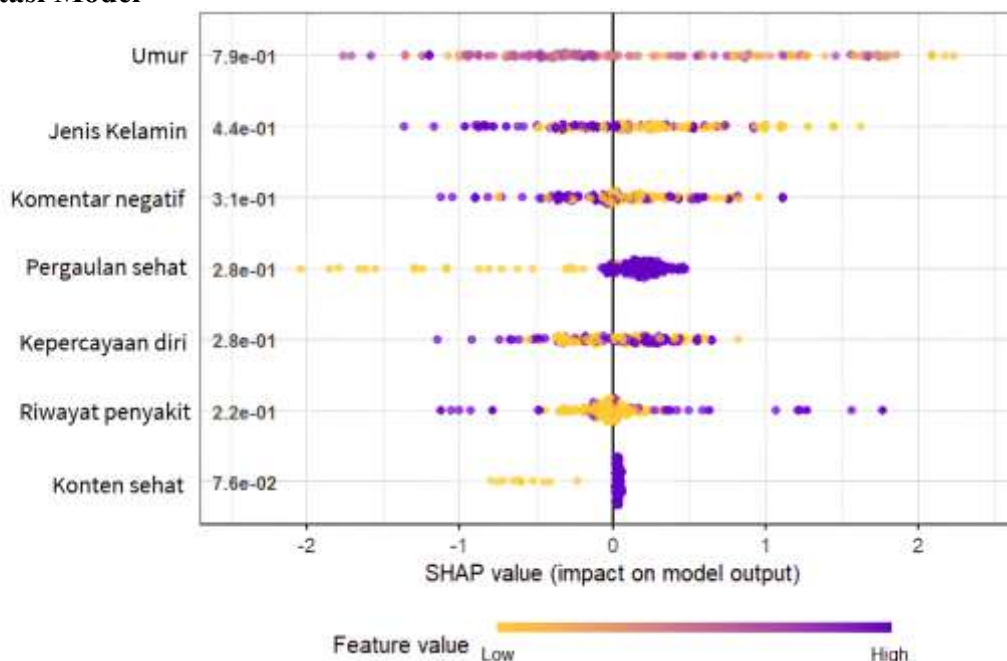
Kategori	Status Diet Pemuda	
	Data latih (85%)	Data uji (15%)
Memulai Diet	119	21
Tidak Memulai Diet	56	9

Model Terbaik

Model	Area Under Curve (AUC)
<i>Naive bayes</i>	0,693
Regresi logistic	0,714
XGBoost	0,725

223

Interpretasi Model



Grafik 5. Nilai SHAP Tiap Variabel Prediktor

Berdasarkan grafik 5, terlihat bahwa empat variabel prediktor utama dalam memprediksi status memulai diet 205 pemuda adalah umur, jenis kelamin, komentar negatif, dan pergaulan sehat. Di antara keempat variabel tersebut, pergaulan sehat memiliki arah nilai SHAP yang paling jelas, sementara arah prediksi dari variabel lain cenderung tidak konsisten. Berdasarkan aspek pergaulan sehat yaitu lingkungan keluarga atau teman sekitar yang sedang menjalani pola hidup sehat berkontribusi terhadap meningkatnya kemungkinan prediksi bahwa pemuda akan memulai diet. Hasil ini sejalan dengan temuan Chung et al. (2017) melalui tinjauan sistematis bahwa perilaku diet remaja dipengaruhi oleh teman sebaya. Meskipun demikian, asosiasi ini bergantung pada jenis kelamin, jenis diet yang dilakukan, serta tingkat kedekatan hubungan pertemanan.

KESIMPULAN

Studi ini bertujuan untuk memprediksi keputusan pemuda di Jakarta pada tahun 2024 untuk memulai diet dengan menggunakan berbagai variabel prediktor melalui pendekatan *machine learning*. Pemuda berusia 16–30 tahun dikategorikan sebagai memulai diet apabila telah menjalani diet selama maksimal tiga bulan. Sementara itu, responden yang telah menjalani diet lebih dari tiga bulan tidak dianggap sebagai unit responden yang valid dalam penelitian. Berdasarkan hasil analisis, model XGBoost menunjukkan performa terbaik dalam memprediksi keputusan pemuda untuk memulai diet dibandingkan model-model lainnya, ditunjukkan oleh nilai AUC yang tertinggi. Nilai SHAP yang positif pada aspek pergaulan sehat menunjukkan bahwa lingkungan keluarga atau teman sekitar yang sedang menjalani pola hidup sehat berkontribusi terhadap meningkatnya kemungkinan bahwa 205 pemuda tersebut akan memulai diet. Namun, karena model yang digunakan bersifat non-inferensial, hubungan lebih lanjut dapat diteliti menggunakan pendekatan inferensial.



REFERENSI

1. Chung, S. J., Ersig, A. L., & McCarthy, A. M. (2017). The Influence of Peers on Diet and Exercise Among Adolescents: A Systematic Review. *Journal of pediatric nursing*, 36, 44–56. <https://doi.org/10.1016/j.pedn.2017.04.010>
2. Davy, S. R., Benes, B. A., & Driskell, J. A. (2006). Sex Differences in Dieting Trends, Eating Habits, and Nutrition Beliefs of a Group of Midwestern College Students. *Journal of the American Dietetic Association*, 106(10), 1673–1677. <https://doi.org/10.1016/j.jada.2006.07.017>
3. Haynos, A. F., Wall, M. M., Chen, C., Wang, S. B., Loth, K., & Neumark-Sztainer, D. (2018). Patterns of weight control behavior persisting beyond young adulthood: Results from a 15-year longitudinal study. *International Journal of Eating Disorders*, 51(9), 1090–1097. <https://doi.org/https://doi.org/10.1002/eat.22963>
4. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning* (2nd ed.). Springer New York, NY. <https://doi.org/https://doi.org/10.1007/978-1-0716-1418-1>
5. Kemenkes. (2015). *Pedoman Umum Pengendalian Obesitas*.
6. Kemenkes. (2019). *Laporan Nasional Riskesdas 2018*. Kementerian Kesehatan Republik Indonesia. <https://repository.kemkes.go.id/book/1323>
7. Kemenkes. (2023a). *Panduan Hari Obesitas Sedunia Tahun 2023*. <https://ayosehat.kemkes.go.id/buku-panduan-hari-obesitas-sedunia-2023>
8. Kemenkes. (2023b). *Survei Kesehatan Indonesia (SKI) 2023 Dalam Angka*.
9. Matera, C., Nerini, A., & Stefanile, C. (2013). The role of peer influence on girls' body dissatisfaction and dieting. *Revue Européenne de Psychologie Appliquée*, 63, 67–74. <https://doi.org/10.1016/j.erap.2012.08.002>
10. Riadi, Muchlisin. (2021). *Fungsi, Aspek dan Jenis-jenis Diet*. <https://www.kajianpustaka.com/2019/08/fungsi-aspek-dan-jenis-jenis-diet.html>
11. Ok, E., & Emmanuel, M. (2025). *XGBoost Model Optimization Techniques*.
12. Oktrisia, C., Prabamurti, P. N., & Shaluhiah, Z. (2021). Beberapa Faktor yang Berhubungan dengan Perilaku Diet Remaja Akhir Usia 18–22 Tahun pada Konsumen Herbalife di Kota Semarang. *Jurnal Kesehatan Masyarakat*, 9(2), 157–165. <https://doi.org/10.14710/jkm.v9i2.29467>
13. Prastiwi, evi dwi, agustina, wiwik, & Fatmawati, diana noor. (2019). Pengaruh Obesitas Sentral Terhadap Status Kesehatan Karyawan. *Jurnal Ilmiah Kesehatan Media Husada*, 8(1), 1–7. <https://doi.org/10.33475/jikmh.v8i1.184>
14. Rodríguez-Pérez, R., & Bajorath, J. (2020). Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design*, 34. <https://doi.org/10.1007/s10822-020-00314-0>
15. Tan, Y. W. B., Lau, J. H., AshaRani, P. V, Roystonn, K., Devi, F., Lee, Y. Y., Whitton, C., Wang, P., Shafie, S., Chang, S., Jeyagurunathan, A., Chua, B. Y., Abdin, E., Sum, C. F., Lee, E. S., & Subramaniam, M. (2022). Dietary patterns of persons with chronic conditions within a multi-ethnic population: results from the nationwide Knowledge, Attitudes and Practices survey on diabetes in Singapore. *Archives of Public Health*, 80(1), 62. <https://doi.org/10.1186/s13690-022-00817-2>



16. World Health Organization. (2000). *Obesity: Preventing and Managing the Global Epidemic. Report of a WHO consultation*. World Health Organization.
17. World Health Organization. (2003). *Diet, Nutrition and the Prevention of Chronic Diseases: Report of a Joint WHO/FAO Expert Consultation, Geneva, 28 January - 1 February 2002*. World Health Organization. <https://iris.who.int/handle/10665/42665>