

ABSTRAK

Obesitas merupakan masalah kesehatan masyarakat yang semakin meningkat dan dipengaruhi oleh berbagai faktor saling berkaitan, seperti pola konsumsi, aktivitas fisik, faktor genetik, kondisi lingkungan, tingkat pendidikan, dan status sosial ekonomi. Kompleksitas faktor tersebut menuntut pendekatan analitik yang mampu menangkap hubungan non-linear antarvariabel. Penelitian ini bertujuan menganalisis faktor-faktor yang berpengaruh terhadap obesitas, mengidentifikasi faktor dominan penyebab obesitas, serta menguji akurasi model prediksi menggunakan algoritma *Random Forest*.

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan *data mining* yang dilaksanakan selama lima bulan (Oktober 2025–Februari 2026). Data primer dikumpulkan melalui kuesioner daring dari 300 responden, dengan dataset memuat 14 atribut demografis dan kebiasaan kesehatan. Tahap *preprocessing* meliputi pembersihan data, perhitungan BMI berdasarkan standar WHO, *label encoding* variabel kategorik, dan seleksi 12 fitur relevan. Data dibagi 80% latih dan 20% uji, lalu dianalisis menggunakan *RandomForestClassifier* dari Scikit-learn.

Hasil evaluasi model menunjukkan akurasi 81,67% dengan nilai presisi, *recall*, dan F1-score rata-rata tertimbang masing-masing sebesar 0,83, 0,82, dan 0,82. Analisis *feature importance* mengungkapkan bahwa Usia (UA) adalah faktor paling dominan dengan skor 0,27, diikuti Kebiasaan Sarapan (KS), Kebiasaan Mengonsumsi Cemilan dan *Snack* (KMCS), dan Aktivitas Fisik Harian (AF). Meskipun korelasi linear antara variabel kebiasaan dan BMI cenderung lemah, variabel-variabel tersebut tetap berkontribusi terhadap perubahan nilai BMI responden. Penelitian ini membuktikan bahwa metode *Random Forest* efektif sebagai pendekatan *machine learning* untuk menganalisis faktor dominan obesitas dan memprediksi kategori BMI berdasarkan data kebiasaan responden.

Kata Kunci: Obesitas, *Random Forest*, *Feature Importance*, Indeks Massa Tubuh (BMI), *Machine Learning*, Faktor Dominan