ABSTRACT

Machine learning is a branch of artificial intelligence that can learn patterns from data to generate automatic predictions and has been widely applied in the health sector, including the detection of obesity risk. This study aims to compare the performance of the Random Forest and C4.5 algorithms in predicting obesity risk using survey data from 446 respondents aged 18–50 years, with lifestyle variables such as fast food consumption, sugary drink intake, fruit and vegetable consumption, exercise habits, sleep duration, smoking status, family history of obesity, and stress levels. The data underwent preprocessing, splitting with a 70:30 ratio, and stratified 5-fold cross validation, then evaluated using accuracy, precision, recall, F1-score, and AUC metrics. The results show that Random Forest outperformed C4.5 with an accuracy of 86.56%, an F1-score of 0.87, and an AUC of 0.984, while C4.5 achieved an accuracy of 85.82%, an F1-score of 0.86, and an AUC of 0.889. Therefore, Random Forest is considered more optimal in predicting obesity risk and has the potential to be used as the basis for developing decision support systems in the healthcare field.

Keywords: Machine Learning, Random Forest, C4.5, Risk Prediction, Obesity

ABSTRAK

Machine learning merupakan salah satu pendekatan kecerdasan buatan yang mampu mempelajari pola data untuk menghasilkan prediksi secara otomatis dan telah banyak dimanfaatkan dalam bidang kesehatan, termasuk untuk mendeteksi risiko obesitas. Penelitian ini bertujuan membandingkan kinerja algoritma *Random Forest* dan *C4.5* dalam memprediksi risiko obesitas menggunakan data survei publik sebanyak 446 responden usia 18–50 tahun dengan variabel gaya hidup seperti konsumsi makanan cepat saji, minuman berpemanis, buah dan sayur, aktivitas olahraga, durasi tidur, kebiasaan merokok, riwayat obesitas keluarga, serta stres. Data diproses melalui tahap preprocessing, pembagian dataset 70:30, dan validasi *stratified* 5-*fold cross validation*, kemudian dievaluasi menggunakan metrik akurasi, presisi, *recall*, *F1-score*, dan AUC. Hasil penelitian menunjukkan *Random Forest* lebih unggul dengan akurasi 86,56%, *F1-score* 0,87, dan *AUC* 0,984, sedangkan *C4.5* memperoleh akurasi 85,82%, *F1-score* 0,86, dan *AUC* 0,889. Dengan demikian, *Random Forest* dinilai lebih optimal dalam memprediksi risiko obesitas dan berpotensi dijadikan dasar pengembangan sistem pendukung keputusan di bidang kesehatan.

Kata Kunci*: Machine Learning, Random Forest, C4.5,* Prediksi Risiko*,* Obesitas