

Sindrom metabolik adalah sekelompok kondisi yang dapat meningkatkan risiko terjadinya penyakit kardiovaskular. Obesitas dan diabetes termasuk ke dalam faktor risiko terjadinya sindrom metabolik pada tubuh seseorang. Langkah pencegahan terjadinya obesitas dan diabetes diperlukan karena seiring bertambahnya tahun prevalensi obesitas semakin meningkat. Saat ini, telah banyak penelitian yang dilakukan terkait penggunaan pembelajaran mesin sebagai alat bantu pengambil keputusan klinis dalam pendeteksian dini yang menunjukkan akurasi yang baik. Namun, kompleksitas model yang membuat proses dalam menghasilkan keluaran diagnosis membuat penggunaan teknologi tersebut belum optimal. Selain itu, dari pihak klinisi merasa bahwa perlu adanya pengawasan manusia dalam hal keputusan yang diambil oleh pembelajaran mesin di lingkungan klinis. Penelitian ini bertujuan untuk mengoptimalkan kemampuan pembelajaran mesin dalam mendeteksi faktor risiko sindrom metabolik dan mengikutsertakan pengetahuan klinisi dalam proses tersebut dengan cara menggabungkan algoritma pembelajaran mesin pengklasifikasi dengan sekumpulan aturan klinis.

Untuk melatih dan menguji model yang dikembangkan pada penelitian ini, digunakan dua jenis dataset, yaitu *dataset* obesitas (2111 data) dan *dataset* diabetes (1879 data). *Dataset* obesitas akan digunakan untuk mengetahui kombinasi model terbaik. Kemudian, model terbaik akan dilatih dan diuji dengan *dataset* diabetes untuk mengetahui performa generalisasi model. Kedua *dataset* tersebut akan dibagi menjadi data pelatihan dan pengujian dengan rasio 70:30. Pada tahap pra-pemrosesan digunakan teknik penskalaan data dengan standardisasi dan normalisasi. Algoritma SVM, Random Forest, dan XGBoost digunakan sebagai algoritma untuk melatih model pengklasifikasi. Sedangkan metode ID3 dan SkopeRules digunakan untuk mengekstraksi sekumpulan aturan. Metode yang digunakan untuk meningkatkan transparansi model pengklasifikasi adalah dengan metode SHAP.

Hasil yang didapatkan menunjukkan bahwa kombinasi algoritma XGBoost dan aturan ID3 adalah kombinasi terbaik karena mendapatkan akurasi tertinggi pada kedua *dataset*, yaitu sebesar 99,37% untuk *dataset* obesitas dan 93,35% pada *dataset* diabetes. Dibandingkan dengan model pengklasifikasi tanpa kombinasi aturan, hasil tersebut menunjukkan peningkatan sebesar 4,26% untuk *dataset* obesitas dan 2,07% untuk *dataset* diabetes. Fitur baru dari implementasi aturan menunjukkan kontribusi yang baik dalam membantu proses pengklasifikasian. Kemudian, metode SHAP yang digunakan berhasil menampilkan grafik kontribusi fitur-fitur pada *dataset* terhadap klasifikasi suatu *instance*.

Kata kunci: obesitas, diabetes, pembelajaran mesin, svm, random forest, xgboost, berbasis aturan, shap

*Metabolic syndrome is a cluster of conditions that can increase the risk of cardiovascular diseases. Obesity and diabetes are among the risk factors that contribute to the development of metabolic syndrome in individuals. Preventive measures against obesity and diabetes are necessary, as the prevalence of obesity has been steadily increasing over the years. Numerous studies have explored using machine learning as a clinical decision-support tool for early detection and showed high accuracy. However, the complexity of the models, which complicates the diagnostic output process, has limited the optimal utilization of this technology. Additionally, clinicians have expressed the need for human oversight in the decisions made by machine learning in clinical settings. This research aims to optimize the capability of machine learning in detecting risk factors for metabolic syndrome while incorporating clinical expertise by integrating classification machine learning algorithms with a set of clinical rules.*

*This study utilized two datasets to train and test the developed model, the obesity dataset (2111 records) and the diabetes dataset (1879 records). The obesity dataset will be used to identify the best model combination, after which the optimal model will be trained and tested using the diabetes dataset to evaluate its generalization performance. Both datasets were split into training and testing sets with a 70:30 ratio. Data scaling techniques, including standardization and normalization, were applied during pre-processing. The algorithms used to train the classification models include SVM, Random Forest, and XGBoost, while the ID3 and SkopeRules methods were used to extract sets of rules. The SHAP method is used to enhance the transparency of the classification model.*

*The results indicate that the combination of the XGBoost algorithm and ID3 rules yielded the best performance, achieving the highest accuracy on both datasets: 99.37% for the obesity dataset and 93.35% for the diabetes dataset. Compared to the classification models without rule combinations, these results represent an improvement of 4.26% for the obesity dataset and 2.07% for the diabetes dataset. The new features from the rule implementation demonstrate a significant contribution in aiding the classification process. Furthermore, the SHAP method successfully displays the contribution graphs of the features in the dataset towards the classification of an instance.*

**Keywords :** *Obesity, Diabetes, Machine Learning, SVM, Random Forest, XGBoost, Rule-Based, SHAP*