

ABSTRAK

Penyalahgunaan narkoba merupakan permasalahan kesehatan masyarakat global yang berdampak signifikan terhadap individu maupun sistem sosial. Deteksi dini terhadap individu berisiko tinggi menjadi langkah strategis dalam upaya pencegahan dan intervensi. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Logistic Regression* dan *Support Vector Machine* dalam memprediksi risiko penggunaan narkoba menggunakan *Drug Consumption (Quantified) Dataset* dari UCI Machine Learning Repository yang berjumlah 1.885 observasi. Proses penelitian meliputi tahap pra-pemrosesan data berupa pembersihan dan normalisasi fitur, seleksi fitur menggunakan *Recursive Feature Elimination*, serta penanganan ketidakseimbangan kelas dengan metode SMOTE. Evaluasi model dirancang menggunakan berbagai metrik klasifikasi seperti *accuracy*, *precision*, *recall*, *F1-score*, AUC-ROC, *Matthews Correlation Coefficient*, dan *G-Mean*, serta pengujian statistik untuk menilai perbedaan performa kedua algoritma. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem skrining dini berbasis *machine learning* sebagai pendukung pengambilan keputusan dalam pencegahan penyalahgunaan narkoba.

Kata Kunci: Penyalahgunaan narkoba, Logistic Regression, Support Vector Machine, Klasifikasi, Machine Learning,

ABSTRACT

Drug abuse constitutes a global public health issue that significantly affects individuals as well as social systems. Early detection of high-risk individuals represents a strategic approach in prevention and intervention efforts. This study aims to compare the performance of the Logistic Regression and Support Vector Machine algorithms in predicting the risk of drug use using the Drug Consumption (Quantified) Dataset from the UCI Machine Learning Repository, consisting of 1,885 observations. The research process includes data preprocessing stages such as data cleaning and feature normalization, feature selection using Recursive Feature Elimination, and handling class imbalance through the SMOTE method. Model evaluation is conducted using various classification metrics, including accuracy, precision, recall, F1-score, AUC-ROC, Matthews Correlation Coefficient, and G-Mean, as well as statistical testing to assess performance differences between the two algorithms. This study is expected to contribute to the development of a machine learning-based early screening system to support decision-making in the prevention of drug abuse.

Keywords: *Drug abuse, Logistic Regression, Support Vector Machine, classification, machine learning*