

ABSTRACT

Chronic kidney disease (CKD) is one of the leading causes of death globally, with rates continuing to rise. It often goes undetected in its early stages due to nonspecific symptoms. The condition becomes more complex when accompanied by comorbidities, making early detection essential to prevent serious complications. This study aims to develop a CKD classification prediction system for both patients with and without comorbidities using the Logistic Regression algorithm, utilizing risk factors extracted from outpatient medical records at Al-Ihsan General Hospital, Bandung, 2024, including age, gender, blood pressure (systolic and diastolic), urea levels, creatinine, hemoglobin, leukocytes, erythrocytes, and blood sugar levels. Model development was conducted following the CRISP-DM methodology, which includes stages from business understanding to implementation into an interactive dashboard. The model was developed through a hyperparameter optimization process using the grid search technique, achieving the highest accuracy of 80% at a 90:10 data split ratio with the optimal configuration: solver: liblinear, penalty: l1, C: 0.01, max_iter: 1000, and class_weight: none. Model performance evaluation showed precision of 0.79, recall of 0.85, and F1-score of 0.82 for class 0 (PGK), and precision of 0.80, recall of 0.72, and F1-score of 0.76 for class 1 (PGK with comorbidities). In addition, an AUC value of 0.80 indicates the model's good discriminatory ability in distinguishing between the two classes. Based on feature importance analysis, seven significant features were identified, namely age, gender, blood urea level (urea), blood creatinine level (creatinine), systolic blood pressure, diastolic blood pressure, and random blood sugar level. These features are clinically important as they are directly related to kidney function, systemic pressure, and the patient's metabolic load. The model was integrated into an interactive dashboard based on Streamlit to facilitate prediction and interpretation of clinical features. The results show that Logistic Regression is effective in supporting early detection of PGK based on clinical data.

Keywords—chronic kidney disease, feature importance, interactive dashboard, logistic regression, performance evaluation.