

ABSTRAKSI

Setyo,Ari, 2025, IMPLEMENTASI KINERJA *ALGORITMA MACHINE LEARNING* REGRESSION TERHADAP DATA DIABETES

Tugas Akhir, Program Sarjana, Program Studi Teknik Informatika. Universitas Muhammadiyah Jember.

Pembimbing: (1) Guruh Wijaya, ST.,M.Kom.; (2) Daryanto, S.Kom., M.Kom

Diabetes melitus adalah penyakit kronis dengan prevalensi global yang meningkat pesat, mempengaruhi sekitar 422 juta orang, terutama di negara-negara berpenghasilan rendah dan menengah. Manajemen diabetes yang efektif memerlukan deteksi dini dan intervensi yang tepat waktu. Penelitian ini bertujuan untuk mengembangkan model prediksi yang akurat untuk diabetes melitus menggunakan *tiga algoritma machine learning: Random Forest*, Regresi Logistik, dan *Decision Tree*. Dataset Pima Indians Diabetes, yang terdiri dari 768 catatan pasien dengan berbagai indikator kesehatan, digunakan untuk pelatihan dan evaluasi model. Analisis data eksploratif menunjukkan korelasi yang signifikan antara kadar glukosa, BMI, usia, dan risiko diabetes. Dataset dibagi menjadi 80% data pelatihan (614 data) dan 20% data pengujian (154 data). Data dinormalisasi menggunakan metode Min-Max Scaler untuk memastikan semua fitur berada pada skala yang sama. Model divalidasi menggunakan metode *cross-validation* dan dievaluasi berdasarkan akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa Regresi Logistik mencapai akurasi tertinggi (75%) dan kinerja yang seimbang dalam mengidentifikasi kasus positif dan negatif

Kata Kunci: Diabetes Melitus, *Machine Learning*, Prediksi Diabetes, *Random Forest*, Regresi Logistik, *Decision Tree*

ABSTRACT

Setyo, Ari, 2025, *IMPLEMENTATION OF MACHINE LEARNING REGRESSION ALGORITHM PERFORMANCE ON DIABETES DATA* . Final Project, Undergraduate Program, Informatics Engineering Study Program. Muhammadiyah University of Jember.

Supervisors: (1) Guruh Wijaya, ST., M.Kom.; (2) Daryanto, S.Kom., M.Kom.

Diabetes mellitus is a chronic disease with a rapidly increasing global prevalence, affecting approximately 422 million people, primarily in low- and middle-income countries. Effective diabetes management requires early detection and timely intervention. This study aims to develop an accurate prediction model for diabetes mellitus using three machine learning algorithms: Random Forest, Logistic Regression, and Decision Tree. The Pima Indians Diabetes dataset, consisting of 768 patient records with various health indicators, was used for model training and evaluation. Exploratory data analysis revealed a significant correlation between glucose levels, BMI, age, and diabetes risk. The dataset was divided into 80% training data (614 records) and 20% testing data (154 records). Data were normalized using the Min-Max Scaler method to ensure all features were on the same scale. The model was validated using cross-validation and evaluated based on accuracy, precision, recall, and F1-score. The results showed that Logistic Regression achieved the highest accuracy (75%) and balanced performance in identifying positive and negative cases.

Keywords: Diabetes Mellitus, Machine Learning, Diabetes Prediction, Random Forest, Logistic Regression, Decision Tree