ABSTRACT

A COMPARISON OF THE ACCURACY LEVELS OF NEURAL NETWORK, SUPPORT VECTOR MACHINE, AND RANDOM FOREST ALGORITHMS IN PREDICTING DIABETES MELLITUS USING MACHINE LEARNING

By

Zidni Robby Rodhiya

Zidnirobby17@gmail.com

Diabetes is often referred to as a silent killer because its symptoms are not immediately apparent and are frequently only discovered after serious complications arise. This contributes to the continuously increasing prevalence of diabetes. The ability to predict an individual's likelihood of developing diabetes quickly and accurately is expected to help reduce its growing prevalence. This study aimed to compare the accuracy levels of Neural Networks, Random Forests, and Support Vector Machine algorithms in predicting diabetes. The data used in this study were obtained from a public dataset on Kaggle, consisting of 100,000 records with nine attributes: gender, age, hypertension, heart disease, smoking history, BMI, HbA1c level, blood glucose level, and diabetes. The preprocessing stages included resampling, normalization, encoding, and data splitting. Evaluation metrics used in this study were accuracy, precision, recall, and F1-score. The analysis concluded that the accuracy values for the Neural Network, Random Forest, and Support Vector Machine algorithms were 88.29%, 90.2%, and 88.76%, respectively. Therefore, the Random Forest algorithm yielded the highest accuracy among the three.

Keywords: Comparison, Prediction, Algorithm, Machine Learning