ABSTRACT

Diabetes is one of the most common chronic diseases affecting millions of people worldwide. Early detection of diabetes is very important to reduce health risks and support timely treatment. This project focuses on building a machine learning-based diabetes prediction system using a Decision Tree Classifier. The dataset is loaded using Python in Google Colab, and preprocessing steps such as checking null values, removing missing data, and eliminating duplicates are performed to improve dataset quality.

The features are separated into input variables and the target output (Outcome), and the dataset is split into training and testing sets using the train-test split method. A Decision Tree model is trained on the training dataset and then tested on unseen data to evaluate its performance. The system measures accuracy using the accuracy score metric.

Additionally, the model is used for real-time prediction by taking user input values like Glucose, Blood Pressure, Insulin, BMI, and Age and predicting whether the patient is affected or not affected by diabetes. This project demonstrates how Decision Tree classification can be applied in healthcare for effective diabetes prediction and supports the development of simple, accurate, and user-friendly diagnostic tools.