Diabetes Prediction System

Abstract

Diabetes is a chronic health condition affecting millions worldwide, and early detection is crucial for effective treatment and management. This project presents a machine learning-based system for predicting the likelihood of diabetes in patients based on various medical attributes. Using the publicly available Pima Indians Diabetes dataset, we preprocess the data by handling missing values and applying feature scaling. Several classification algorithms - including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine - are implemented and evaluated to identify the most accurate model. Performance metrics such as accuracy, precision, recall, and ROC-AUC score are used to assess the models. The best-performing model is then saved and optionally deployed using a web interface, enabling users to input patient data and receive real-time predictions. This system can serve as a decision-support tool for medical professionals and help in early diagnosis, thereby improving patient outcomes.