Diabetes Prediction Using Machine Learning

Suvradeep Sahana, Suvankar Biswas, Souhardya Sarkar
Debadrita Aich Sarkar, , Sujata Kundu
Department of Information Technology
Narula Institute of Technology, Kolkata, India
suvradeepsahana10@gmail.com, sb7679250005@gmail.com, sarkarsouhardya67@gmail.com,
debadritaaichsarkar42@gmail.com, sujata.kundu@nit.ac.in

Abstract: *Diabetes is a persistent metabolic condition that impacts millions of individuals globally. Timely identification and intervention are essential for effective management and the prevention of associated complications. This project explores the viability of utilizing machine learning algorithms to forecast the onset of diabetes in individuals. The research employs the well- known Pima Indians Diabetes Dataset, which includes a diverse array of physiological and demographic characteristics. A thorough analysis was carried out using various classification algorithms, such as Logistic Regression, Support Vector Machines (SVM), Decision Trees, Random Forest, and K-Nearest Neighbors (KNN). Prior to training the models, extensive data preprocessing methods were applied to handle missing values and standardize features, thereby ensuring optimal performance and generalizability of the models. The evaluation of the models was conducted using a range of metrics, including accuracy, precision, recall, F1-score, and the area under the Receiver Operating Characteristic curve (AUC). The findings indicate that machine learning has significant potential in accurately predicting diabetes, with some algorithms demonstrating superior performance over others. This study underscores the promise of machine learning as an essential resource for healthcare professionals in the early detection of diabetes, facilitating proactive interventions and ultimately enhancing patient outcomes*

Keywords: Diabetes.