

Diabetes Prediction Using Machine Learning

Introduction

Diabetes is a growing chronic disease worldwide, with millions of people at risk of the potentially deadly complications of their growing heart disease as well as kidney failure and blindness when therapy is ignored. Early detection of medical conditions not only helps to reduce risks but also helps the patient in recovering from the illness sooner. In this project, we build a machine learning model that will take the medical data of people and tell how much risk of diabetes they have.

Problem Statement

The existing methods for detecting diabetes depend mainly on clinical testing that shows high costs and requires time and presents access barriers for numerous people. The utilization of historical medical data for machine learning model training creates an effective system that predicts diabetes risk both accurately and cost-efficiently on a large scale.

Objectives

The primary objectives of this project are:

- The development of a machine learning model should utilize key medical features including glucose level, BMI, and age to predict diabetes outcomes.
- Evaluate the performance of different ML algorithms and identify the most accurate model.

Methodology

Data Collection

The project will utilize the PIMA Indians Diabetes Dataset from the UCI Machine Learning Repository, which contains 768 records of female patients along with their medical attributes.

Data Preprocessing

Missing values or zero values in key attributes will be replaced with median values.

All variables need standardization because it makes features uniform regarding their measurement units.

Machine Learning Models

The following machine-learning algorithms will be implemented and evaluated:

- Logistic Regression
- Random Forest
- Support Vector Machine (SVM)
- Gradient Boosting
- ANN (if Possible)

Model Training and Evaluation

- Split the dataset into training and testing sets.
- Train the models and evaluate their performance using accuracy, precision, recall, F1-score, and ROC-AUC metrics.
- Perform hyperparameter tuning to optimize the best-performing model.

Expected Outcomes

The research achieves high predictive accuracy for diabetes using machine learning models.

Insights into the importance of various medical attributes in diabetes prediction.

Conclusion

The project uses machine learning to develop a dependable and efficient method for accessible diabetes prediction. The system enables timely detection which both enhances treatment results and cuts down medical expenses.