VIT BHOPAL UNIVERSITY

Vityarthi Project

PROJECT REPORT

On

DIABETES RISK PREDICTION USING MACHINE LEARNING

Submitted by:

Sk Mohammd Zunaid (25BHI10085)

B.Tech CSE (Health Informatics)

Submitted to:

Faculty Name: _____Monika Vyas_____

Abstract

This project focuses on predicting diabetes risk using machine learning. The Pima Indians Diabetes Dataset is used to train a Logistic Regression model. The final system predicts whether a person is at high or low risk of diabetes.

Introduction

Diabetes is a global health concern requiring early detection to prevent complications. Machine learning enables predictive analytics to identify individuals at risk. This project aims to implement an end-to-end ML workflow for diabetes prediction.

Objectives

- Build a classification model to predict diabetes risk.

- Perform data preprocessing and scaling.

- Train and evaluate a logistic regression model.

- Implement a CLI prediction system.

- Document and present the complete workflow.

Dataset Description

The Pima Indians Diabetes Dataset includes 768 records with 8 medical features such as glucose level, BMI, blood pressure, and insulin. The target variable (Outcome) indicates diabetes diagnosis (1) or non-diabetic (0).

Methodology

1. Data Loading

2. Preprocessing & Missing Value Handling

3. Feature Scaling using StandardScaler

4. Train-Test Split (80-20)

5. Model Training using Logistic Regression

6. Model Evaluation using accuracy, precision, recall, and confusion matrix

7. Deployment using a CLI-based predictor

Results

Conclusion

The project successfully demonstrates an end-to-end machine learning solution for diabetes prediction. Future improvements include integrating more ML models, adding a Streamlit web interface, and performing hyperparameter tuning.

Future Enhancements

- Add RandomForest, SVM, and XGBoost models

- Deploy using Streamlit or Flask

- Integrate SHAP for feature explainability

- Build mobile-friendly UI