Diabetes Prediction and Risk Analysis

Abdul Qadeer (2022-CS-835)
Supervised by: Mr. Asim Naveed

Department of Computer Science
University of Engineering and Technology, Lahore
Faisalabad Campus

Monday, December 23, 2024

Abstract

This project aims to develop a real-time diabetes prediction system that forecasts the likelihood of diabetes based on key health parameters such as age, BMI, HbA1c level, blood glucose level, hypertension, heart disease, gender, and smoking history. By integrating machine learning models, the system provides accurate predictions and risk analysis, which can be utilized for preventive healthcare measures.

1 Introduction

1.1 Background

Diabetes is a chronic disease that affects millions of people worldwide. Early prediction and intervention can significantly reduce the risk of complications. Traditional methods of diabetes prediction often rely on statistical models and historical data, which may not be effective in real-time scenarios. Advances in machine learning provide an opportunity to create more accurate and adaptive prediction systems.

1.2 Objective

The objective of this project is to develop a diabetes prediction system using machine learning techniques. The system aims to forecast the likelihood of diabetes based on various health parameters and provide users with actionable insights.

1.3 Scope

This project focuses on predicting diabetes using real-time health data. It aims to:

- Utilize machine learning models for prediction.
- Provide an interactive user interface for data input and results display.
- Offer risk analysis based on user inputs.

2 Methodology

2.1 Data Collection

The dataset used for training the models was collected from various health records and includes features such as age, BMI, HbA1c level, blood glucose level, hypertension, heart disease, gender, and smoking history.

2.2 Data Preprocessing

The data was cleaned by removing duplicates and handling missing values. One-hot encoding was applied to categorical variables to prepare the data for machine learning algorithms.

2.3 Model Development

Multiple machine learning models were developed, including:

- Random Forest Classifier
- Decision Tree Classifier
- Logistic Regression
- Voting Classifier (ensemble of the above models)

The models were trained using a training dataset and evaluated on a separate test dataset.

2.4 User Interface

The user interface was developed using Streamlit, allowing users to input their health data and receive predictions and risk analysis in real-time.

3 Functional and Non-Functional Requirements

3.1 Functional Requirements

- The system must fetch real-time health data from user inputs.
- The system must predict the likelihood of diabetes based on the input data using trained machine learning models.
- The system must provide risk analysis and feedback based on the predictions.
- The system must generate visualizations to represent risk levels and probabilities.

3.2 Non-Functional Requirements

- **Performance:** The prediction must be generated within 2 seconds after the data is retrieved.
- **Usability:** The user interface must be intuitive and easy to navigate for users of all technical backgrounds.
- **Scalability:** The system should handle multiple user requests simultaneously without degrading performance.
- **Reliability:** The system must provide accurate predictions with a high degree of confidence.

4 Results

The preliminary results indicate that the system can accurately predict diabetes with a high degree of accuracy. The Voting Classifier achieved the best performance among the models tested. The system provides users with a risk analysis based on their input data, helping them understand their health status better.

4.1 Visualizations

The system generates various visualizations, including:

- Risk breakdown bar chart
- Risk probability graphs based on BMI, HbA1c level, and blood glucose level

5 Conclusion

The diabetes prediction system developed in this project demonstrates the potential of machine learning in healthcare. By providing real-time predictions and risk analysis, the system can assist individuals in making informed health decisions.

6 References

- Diabetes Prediction Dataset: `https://example.com/diabetes_dataset`
- Machine Learning Techniques: `https://example.com/machine_learning`
- Streamlit Documentation: `https://docs.streamlit.io`