University of Essex

PILOT STUDY PROPOSAL

CE802 Machine Learning

Prepared For :

Healthcare provider

MAHEN92802

Naresh Mahendiran (MSc in Artificial intelligence)

Highlights

- Detection of the risk factors of developing diabetes using machine learning (ML)
- Prediction of developing diabetic using machine learning (ML) algorithms.
- Justification of the proposed ML-based system

Abstract

The purpose of this pilot project is to evaluate the feasibility of identifying individuals who are at high risk of developing diabetes by using machine learning techniques in electronic medical data. By categorizing people as either having a high risk of developing diabetes or a low risk of doing so, we propose a binary classification. Age, sex, BMI, family history of diabetes, blood pressure, fasting blood glucose levels, and HbA1C values will all be used to estimate the risk of developing diabetes. Decision trees classifier is the machine learning techniques that we recommend using because of their interpretability, ability to handle both categorical and continuous data, and effectiveness in previous studies. To evaluate the system's performance, we'll utilize cross-validation, accuracy, precision, recall, F1-score, and AUC-ROC.This pilot experiment will pave the way for larger-scale investigations and provide insight into the possibility of detecting patients at risk of developing diabetes using machine learning.

Aims

The purpose of this pilot study is to determine whether it may be feasible to identify individuals who have a higher risk of developing diabetes by utilizing machine learning techniques.

Methods

We will use electronic medical records to compile data on those who are at risk of developing diabetes. The data will be preprocessed to address missing values and outliers, and feature selection will be done to identify the most illuminating features. We will train a decision tree model based on the selected traits, assess the model's performance using cross-validation, and then examine the model to determine the main features and their associations. By providing vital data on the possibility of utilizing machine learning to identify patients at risk of contracting diabetes, this pilot project will pave the way for more substantial research.

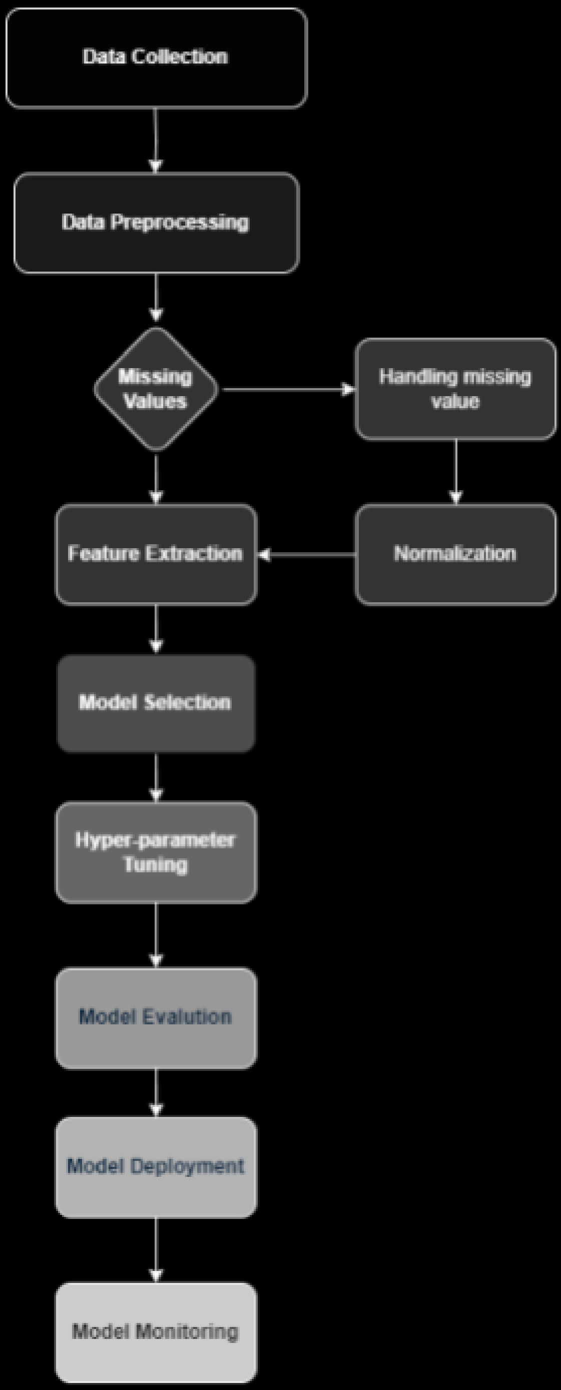
Conclusions

We propose a comprehensive approach for identifying individuals who are at a high risk of developing diabetes using electronic medical records. We will undertake feature selection, data preprocessing, model training, performance evaluation, and result interpretation in order to comprehend the model's behavior. By demonstrating the potential of using machine learning to identify people at risk of developing diabetes, this pilot project will open the door for further investigation.

Introduction

This proposal tests the hypothesis that individuals with a high risk of developing diabetes can be diagnosed using machine learning methods. By identifying patients at risk, the healthcare practitioner can take preventative measures to reduce the likelihood of developing diabetes.

Machine learning system pipeline :



Informative Features

Some significant characteristics that can be used to predict the risk of developing diabetes include age, sex, body mass index (BMI), family history of diabetes, blood pressure, fasting blood glucose levels, and hemoglobin A1C (HbA1C) readings. These traits have been found to significantly increase the risk of developing diabetes.

Learning Procedure:

Some machine learning techniques that can be applied to this problem include decision trees, logistic regression, random forests, and support vector machines (SVMs). Among these methods, we suggest using decision trees since they are easy to grasp, can handle both continuous and categorical data, and have a track record of success.

Evaluation

To evaluate the effectiveness of the system, we'll utilize cross-validation, which separates the dataset into training and testing sets. The decision tree model will be trained using the training set, and its performance will be evaluated on the testing set. We will also use evaluation metrics like accuracy, precision, recall, and F1-score to assess the model's efficacy. Additionally, we will use the area under the receiver operating characteristic curve (AUC-ROC) to evaluate the model's performance.

Conclusion

In conclusion, the purpose of this pilot project is to determine whether it is feasible to identify patients who have a high risk of acquiring diabetes using machine learning. The suggested predictive task is classification, and the suggested learning method is SVMs. The predictive model will be trained using a range of useful characteristics, and the performance of the model will be assessed using a number of measures, including AUC-ROC. The findings of this study will shed important light on how machine learning may be used to enhance healthcare outcomes.

About Us

Naresh Mahendiran (MAHEN92802)
MSc in Artificial Intelligent
University of Essex
School of Computer Science and Electronic Engineering

As a self-employed Machine Learning consultant, I provide firms wishing to integrate data analytics into their operations scientific advice and consulting services.

Thank you, and we look forward to working with you.

Reference:

- Prediction of progression from pre-diabetes to diabetes: Development and validation of a machine learning model :https://onlinelibrary.wiley.com/doi/full/10.1002/dmrr.3252
- A survey on diabetes risk prediction using machine learning approaches : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10041290/