**Report**

**Diabetes Prediction**

Abstract

* Diabetes is one of deadliest diseases in the world. It is not only a disease but also a creator of different kinds of diseases like heart attack, blindness, kidney diseases, etc.
* In this project we will**predict** and perform evaluation Metrics of the**Early Stage Diabetes** Risk**Prediction** dataset. The dataset can be found on kaggle.

Introduction

* Diabetes is chronic disease which is measured by blood sugar levels
* It has became a severe threat to peoples lives
* It leads to serious damage to the heart, liver, kidneys and nerves
* Although early prediction and diagnosis of diabetes can prevent some serious damage over the human body
* With Machine learning models we can predict early stages of diabetes

Proposed method with Architecture

Classification is one of the most important decision making techniques in many real world problem. In this work, the main objective is to classify the data as diabetic or non-diabetic and improve the classification accuracy. The main objective of our model is to achieve high accuracy. Classification accuracy can be increase if we use much of the data set for training and few data sets for testing. Thus, it is observed that techniques like Support Vector Machine, Logistic Regression, and Random Forest are most suitable for implementing the Diabetes prediction system

Methodology

**Dataset :** for this project we used ‘The Early Stage Diabetes Risk Prediction’ dataset present in kaggle.

The above dataset contains 16 features, one target column with two class labels ‘Positive’ and ‘Negative’ and 520 rows.

**Method :** Some simple Machine Learning classification models like Logistic Regression, Support Vector Machine, Random Forest classifiers were used to predict the classes.

Among the above used methods the model with high accuracy is taken into consideration.

Implementation

1. Importing Libraries

Key libraries used:

* + - Scikit-learn
    - Pandas, numpy , matplotlib

2. Data Pre-processing

* + Normalization
  + Splitting train and test batches

3. Training a model

4. Prediction

Conclusion

* Machine learning has the great ability to revolutionize the diabetes risk prediction with the help of advanced computational methods and availability of large amount of epidemiological and genetic diabetes risk dataset. Detection of diabetes in its early stages is the key for treatment. This work has described a machine learning approach to predicting diabetes levels. The technique may also help researchers to develop an accurate and effective tool that will reach at the table of clinicians to help them make better decision about the disease status