# Diabetes-Predication Using

# Machine Learning

By

Sai Kumar Dacharla

**CONTENTS**

|  |  |
| --- | --- |
| **Heading** | **Page No** |
| DIABETES PREDICTION USING MACHINE LEARNING | 1 |
| INTRODUCTION | 3 |
| ABSTRACT | 3 |
| DATA SET | 4 |
| APPROACHES TO SOLVE | 5 |
| RESULTS | 6 |
| CONCLUSION | 6 |
| REFERENCES | 7 |

**INTRODUCTION**

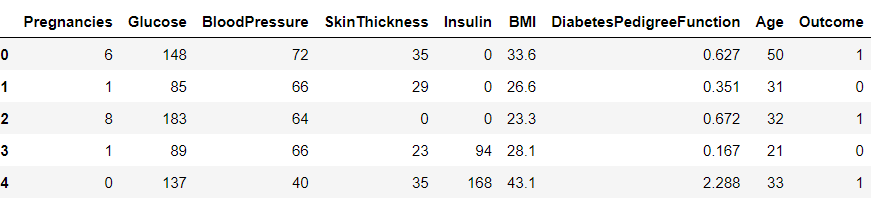
Diabetes prediction using machine learning is an approach to predicting the likelihood of an individual developing diabetes using clinical and demographic data. By training computer algorithms on large datasets, machine learning techniques can identify patterns and relationships in the data that can be used to make accurate predictions. Early detection of diabetes is crucial for effective management of the disease and its associated complications. The use of machine learning in diabetes prediction has the potential to revolutionize the way diabetes is diagnosed and managed by enabling targeted interventions and lifestyle modifications for high-risk individuals. Overall,diabetes prediction using machine learning is an exciting area of research with significant potential to improve public health.

**ABSTRACT**

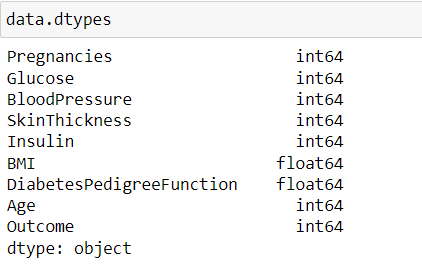
In this report, we will try to understand the factors that are contributing to the mental health of a person. The dataset used in this project is from Kaggle that measures attitudes towards diabetes. The use of machine learning in diabetes prediction has the potential to revolutionize the way diabetes is diagnosed and managed.

**DATA SET**

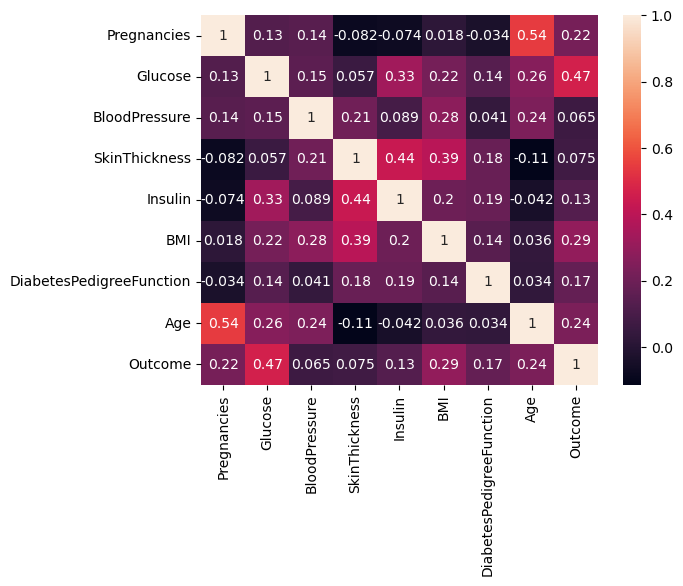
The datasets consists of several medical predictor variables and one target variable, Outcome. Predictor variables like Pregnancies, Glucose, Blood pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age and Outcome.

****

The data types used first are of different types. But I have changed all of them to integers except for the column BMI and Diabetes Pedigree Function.



After performing some exploratory data analysis I have performed correlation for columns to know the dependency of the target column.



**APPROACHES TO SOLVE**

There are many approaches available to do this problem, but I have selected most popular models that are listed below.

Logistic Regression

KNearestNeighbour

Support Vector Machine

Decision Tree Classifier

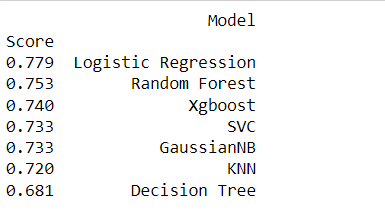
GaussianNB

Random Forest Classifier

XG Boosting

**RESULTS**

From the below table we can clear say that Logistic Regression algorithm have classified almost all the samples correctly expect a few of them.There fore, we can conclude Logistic Regression algorithm is the best among all the other algorithms.



**CONCLUSION**

In conclusion, diabetes prediction using machine learning is a promising approach to identifying individuals at high risk of developing diabetes. By leveraging large datasets and powerful algorithms, machine learning techniques can identify patterns and relationships in the data that can be used to make accurate predictions. Early detection of diabetes is crucial for effective management of the disease, and the use of machine learning in diabetes prediction has the potential to revolutionize the way diabetes is diagnosed and managed. By providing targeted interventions and lifestyle modifications for high-risk individuals, the incidence and complications of diabetes can be reduced. Overall, diabetes prediction using machine learning is an exciting area of research with significant potential to improve public health.

**REFERENCES**

<https://ieeexplore.ieee.org/document/9441935>

<https://iopscience.iop.org/article/10.1088/1742-6596/1684/1/012062/pdf>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9018179/>