IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11551

# DIABETES PREDICTION USINGMACHINE LEARNING

**Dr Lokesh S[1], Mr. Channaveeraya W M[2],**

**Alson Chris Dsouza[3], Sandeep[4], Sanjana Ajit Bhat[5], Sanjay S[6]**

Associate Professor, Computer Science and Engineering, NIE, Mysuru, India[1]

Assistant Professor, Computer Science and Engineering, NIE, Mysuru, India[2]

Student, Computer Science and Engineering, NIE, Mysore, India[3-6]

**Abstract:** Too much sugar in the body is a symptom of diabetes, a disease that is very dangerous if left uncontrolled. Diabetes complications: It can occur in many forms, including heart problems, kidney failure, high blood pressure, visual impairment, and organ damage. Early diagnosis of diabetes is important for effective management; therefore, the project will focus on using machine learning (ML) technology for accurate prediction. Using various ML classification and integration methods such as support vector machine (SVM), K nearest neighbor (KNN), decision tree (DT), logistic regression (LR), random forest (RF), and gradient boosting (GB) in the collection. We aim to increase the accuracy of facts from the patient's dataset. Each method showed different accuracy, and Random Forest emerged as the best model in our study. Analysis of the most accurate model demonstrates its ability to accurately predict blood sugar levels, leading to major advances in healthcare predictions.

## I. INTRODUCTION

Diabetes is a chronic disease that affects millions of people worldwide and has a major impact on society, health, and society. The disease affects the control of diabetes due to inadequate production or use of insulin, an important hormone that controls blood sugar. According to the World Health Organization (WHO), 422 million people worldwide have diabetes, mostly in low- and middle-income countries. By 2030, this number is expected to increase to approximately 490 million. Timely detection and diagnosis are important for improving patient outcomes and effective diabetes management. Undiagnosed or untreated diabetes can lead to increased hunger, thirst, and urination due to elevated blood sugar levels. Early diagnosis is key to preventing these problems, and new technologies can aid in this effort. risk of diseases such as heart disease, kidney failure, and vision problems. Machine learning is a branch of artificial intelligence that is changing healthcare by improving disease, diagnosis, and treatment. With the help of machine learning algorithms, doctors can accurately diagnose conditions and develop personalized treatment plans. Research shows that machine learning algorithms outperform statistical models in creating predictive models for many situations. By training models on big data, we aim to uncover important insights for predicting the onset of diabetes in the future. A preliminary phase involves cleaning the data handling missing values, and then searching the data to identify key features. We performed the prediction using various machine learning methods and the Pima Indian Diabetes dataset. Although there are many ways to create good results, it is still difficult to choose the best way, which encourages us to explore popular distribution and integration. The main aim of our study is to develop reliable and accurate prediction models to identify individuals at risk of diabetes. Using advanced machine learning technology and comprehensive data that includes clinical, lifestyle, and demographic variables, we strive to create tools that help doctors make informed decisions, allocate time, and provide preventive care.

## II. LITERATURE REVIEW

This area looks at related investigations on diabetes expectations and the application of machine learning methods in this domain.Relevant ponders incorporate the work of Bir Hussain, David Britain, and Nonso Nnamoko (2018), who displayed their approach of utilizing gathering directed learning to anticipate the onset of diabetes. Their investigation illustrated that the proposed strategy may accomplish more precise forecasts of diabetes onset compared to other techniques.

Another eminent commitment is the paper "Diabetes Forecast Utilizing Machine Learning Calculations and Philosophy" by H. E. Massari, Z. Sabouri, S. Mhammedi, and N. Gherabi, distributed within the Diary of ICT Standardization in 2022. This ponder investigated the utilization of machine learning calculations and philosophy for diabetes prediction.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11551

Ahamed, B. S., Arya, M. S., Sangeetha, S. K. B., & Nancy, V. A. O. (2022) too made noteworthy commitments to the field with their investigations on diabetes mellitus infection forecast and sort classification including prescient modelling using machine learning techniques and classifiers, distributed within the diary Connected Computational Insights and Delicate Computing.

Additionally, Mitushi Soni and Dr. Sunita Varma (2020) displayed their work on diabetes forecasting utilizing machine learning within the Universal Diary of Designing Inquire about Innovation (IJERT).

These articles offer important experiences in the advance and execution of diabetes forecasts and highlight the vital part of machine learning methods in guaranteeing exact and solid expectations.

## III. PROPOSED METHODOLOGY

This section describes the approach we plan to take using various machine-learning techniques to develop a system for predicting diabetes. The various stages involved in the search process are described below.

1.Dataset Acquisition

In this experiment, we used the well-known Pima Indian dataset, which is a public data frequently used in machine learning studies. The database contains information on approximately 768 patients,268 of whom were diagnosed with diabetes. It includes eight important factors: age, body mass index (BMI), skin thickness, blood pressure, diabetes, pregnancy history, insulin levels, and diabetes outcomes.

2.Data Preprocessing

Data preprocessing is an important step, especially in medical data, where missing values and inconsistencies can affect the quality and validity of the data. Solving these issues can improve the accuracy and quality of data collected for analysis. Proper data preparation is essential to obtain accurate results and reliable predictions when using machine learning techniques. Outliers will not be zero in the data set, especially in variables such as skin thickness and body weight index. We correct these issues by replacing zero values with appropriate averages to maintain data accuracy and completeness. The data is divided into training and testing sets. The training method is used to train the algorithm and the model is learned from the features and patterns in the training data. Testing methods are reserved for evaluating the effectiveness of training models.

3.Application of Machine Learning Algorithms

We are working on various machine learning methods, including various classification and integration methods, to predict diabetes using prepared data. Our main goal is to evaluate the effectiveness of these methods and determine their accuracy in predicting blood sugar. We also aim to identify key features that influence the forecasting process. Random forests are known for their high accuracy and ability to handle large amounts of data. It is also an excellent learning method valued for its ability to improve the decision tree by reducing variance. During training, a random forest creates a large number of decision trees and determines the final classification by averaging (regression) or selecting the class working model of a person tree (classification).

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11551

**Fig. 1:** Dashboard of the webpage

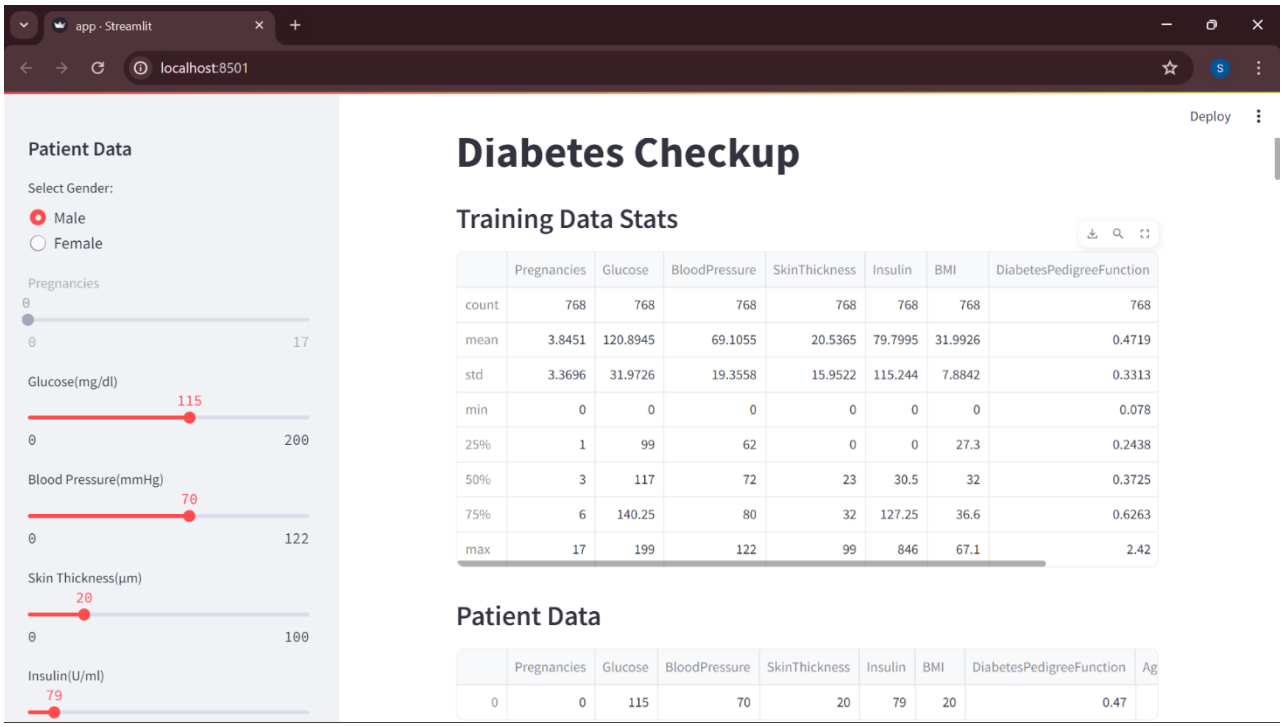
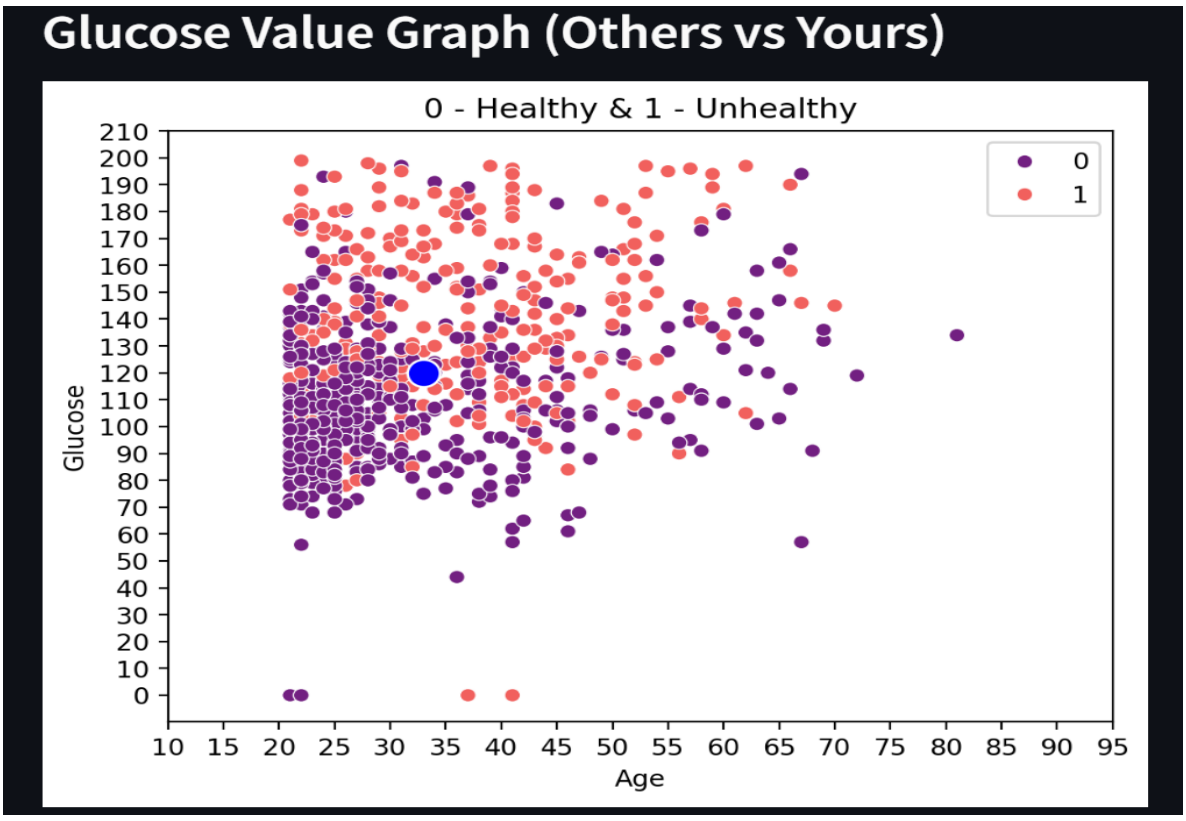



**Fig. 2:** Graphical representation of values



IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11551

**Fig. 3:** Patient data entry to the system


## Training Data Stats

| | Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI | DiabetesPedigreeFunction |
|---|---|---|---|---|---|---|---|
| count | 768 | 768 | 768 | 768 | 768 | 768 | 768 |
| mean | 3.8451 | 120.8945 | 69.1055 | 20.5365 | 79.7995 | 31.9926 | 0.4719 |
| std | 3.3696 | 31.9726 | 19.3558 | 15.9522 | 115.244 | 7.8842 | 0.3313 |
| min | 0 | 0 | 0 | 0 | 0 | 0 | 0.078 |
| 25% | 1 | 99 | 62 | 0 | 0 | 27.3 | 0.2438 |
| 50% | 3 | 117 | 72 | 23 | 30.5 | 32 | 0.3725 |
| 75% | 6 | 140.25 | 80 | 32 | 127.25 | 36.6 | 0.6263 |
| max | 17 | 199 | 122 | 99 | 846 | 67.1 | 2.42 |

## IV. CONCLUSION

Our study involved evaluating various machine learning methods for measuring diabetes. We then created a web application that provides users with access to medical information to predict the risk of diabetes. This has the potential to help with early diagnosis of diabetes and support treatment decisions. We have improved usability by using larger and more diverse data, integrating real-time monitoring, enhancing understanding, and improving interpretation. These advancements are designed to increase the accuracy, validity, and effectiveness of predictive models, ultimately benefiting physicians and patients for better blood transfusion management.

## V. REFERENCES

[1]. Mitushi Soni, Dr. Sunita Varma, 2020, Diabetes Prediction using Machine Learning Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 09 (September 2020).

[2]. H. E. Massari, Z. Sabouri, S. Mhammedi and N. Gherabi, "Diabetes Prediction Using Machine Learning Algorithms and Ontology," in Journal of ICT Standardization, vol. 10, no. 2, pp. 319-337, 202.

[3]. Wang, Y., Hou, R., Ni, B., Jiang, Y., & Zhang, Y. (2023). Development and validation of a prediction model based on machine learning algorithms for predicting the risk of heart failure in middle-aged and older US people with prediabetes or diabetes. Clinical Cardiology, 46(10), 1234-1243.

[4]. Yuvaraj, N., & SriPreethaa, K. R. (2019). Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster. Cluster Computing, Suppl.1, 22, 1-9.

[5]. Doreswamy, G. S., Santosh, K. J., & Nandish. (2018). PREDICTION ACCURACY COMPARISON OF PREDICTIVE MODELS USING MACHINE LEARNING FOR DIABETES DATA SET. International Journal of Advanced Research in Computer Science, 9, 86-88