Mini Project

on

" DIABETES PREDICTION: WEB-BASED APPLICATION FOR EARLY DETECTION AND HEALTH MONITORING"

Submitted to

Sant Gadge Baba Amravati University, Amravati
in partial fulfillment of the requirement for the award of degree of
Bachelor of Engineering in Artificial Intelligence and Data Science

Submitted by

Mr. Adarsh Lilhare

Miss. Arati Paraskar

Mr. Bhupesh Wankhade

Mr. Harshal Pawar

Mr. Shreyash Nangirwar

Under the guidance of

Prof. L. S. Bhattad

Department of Artificial Intelligence and Data Science

P. R. Pote Patil College of Engineering & Management, Amravati

2024-25

CERTIFICATE

This is to certify that the project entitled

“Diabetes Prediction: Web-Based Application for Early Detection and Health Monitoring”

Is a bonafide work and it is submitted to the

Sant Gadge Baba Amravati University, Amravati.

Submitted by

Mr. Adarsh Lilhare

Miss. Arati Paraskar

Mr. Bhupesh Wankhade

Mr. Harshal Pawar

Mr. Shreyash Nangirwar

In the partial fulfillment of the requirement for the award of the degree of Bachelor of Engineering in Artificial Intelligence and Data Science during the academic year of 2024– 25

Prof. L. S. Bhattad	**Prof. Y. M. Babnekar**	**Dr. A. B. Gadicha**
Guide	Project Coordinator	HOD, AI &DS

Department of Artificial Intelligence and Data Science

P. R. Pote (Patil) College of Engineering & Management, Amravati

2024-25

INDEX

DECLARATION

I hereby declare that the Project entitled, **"Diabetes Prediction: Web-Based Application for Early Detection and Health Monitoring"** was carried out and written by me under the guidance of **Prof. L. S. Bhattad**, Dept. of Artificial Intelligence and Data Science, P. R. Pote Patil College of Engineering & Management, Amravati. This work has not been previously formed the basis for the award of any degree or diploma or certificate, nor has it been submitted elsewhere for the award of any degree or diploma.

Place: Amravati

Date:

Mr. Adarsh Lilhare

Miss. Arati Paraskar

Mr. Bhupesh Wankhade

Mr. Harshal Pawar

Mr. Shreyash Nangirwar

Third Year (AI &DS)

ACKNOWLEDGEMENT

It is my utmost duty and desire to express acknowledgement to the various torchbearers who have rendered valuable guidance during the preparation of my Project. First of all, I extend my deepest gratitude to the respected Principal, **Dr. P. M. Jawandhiya,** without whose support, my Project could not have been transformed into its present form.

I am grateful to **Dr. A. B. Gadicha,** Head, Artificial Intelligence and Data Science Department, and my guide, **Prof. L. S. Bhattad,** for providing immense support and guidance. I am beholden for guiding me at every step in the Project. He has honestly guided me throughout, never leaving me unanswered for any of my doubts. It was his constant persuasion, encouragement, inspiration, and able guidance that helped me in completing my Project successfully.

Mr. Adarsh Lilhare

Miss. Arati Paraskar

Mr. Bhupesh Wankhade

Mr. Harshal Pawar

Mr. Shreyash Nangirwar

Third Year (AI &DS)

LIST OF FIGURE

LIST OF ABBREVIATION

Sr.No.	Title	Full form
1	AI	Artificial Intelligence
2	ML	Machine Learning
3	BMI	Body Mass Index

ABSTRACT

Diabetes is a chronic metabolic disorder that affects millions of people globally. It is primarily characterized by elevated levels of blood glucose, which, if left unmanaged, can lead to severe complications such as cardiovascular diseases, kidney failure, nerve damage, and vision problems. Early diagnosis and continuous monitoring are crucial for effective management and prevention of these health issues.

This project focuses on leveraging machine learning regression techniques to improve the prediction and diagnosis of diabetes. By utilizing the PIMA Indian Diabetes Dataset, which includes various medical and personal health-related attributes (such as age, BMI, glucose levels, insulin levels, and more), the system is able to learn and recognize complex patterns associated with the onset of diabetes.

Through this data-driven approach, the model is trained to identify key risk factors and predict the likelihood of a patient developing diabetes. Unlike traditional diagnostic methods that may be time-consuming or require expensive lab tests, this solution offers a faster and more accessible way to screen individuals, especially in remote or underserved regions.

To make this solution practical and user-friendly, the machine learning model is integrated into a web-based application. This application allows users—including patients, healthcare providers, and researchers—to conveniently enter personal health data through an intuitive interface. The system processes the input in real-time, provides predictive insights, and displays the results in an interactive and easy-to-understand manner. Visualizations and risk indicators further help users interpret the results and take appropriate actions.

Overall, the project aims to empower early detection and proactive healthcare management by combining data science with an accessible digital platform. It supports better health outcomes and has the potential to be integrated into broader health screening programs or mobile health solutions in the future.

1. Introduction

Diabetes mellitus is one of the most common chronic diseases worldwide. According to the World Health Organization (WHO), the global prevalence of diabetes has nearly doubled in recent decades. This metabolic disorder leads to high blood sugar levels over an extended period and is caused by either the pancreas not producing enough insulin or the body's cells failing to respond properly to insulin. The long-term effects of diabetes can include cardiovascular disease, kidney failure, eye damage, and nerve damage.

Early prediction of diabetes is essential for better disease management and potentially preventing serious health complications. Traditional diagnostic methods can often be invasive and costly, which makes the use of non-invasive, cost-effective machine learning techniques an appealing approach for predicting diabetes at an early stage.

The goal of this project is to develop a web-based machine learning model that can accurately predict diabetes based on key clinical parameters. The system will offer users a straightforward interface to input their health information and receive an instant prediction of their diabetic status.

1.1 Motivation

In today’s digital landscape, the integration of artificial intelligence (AI) in healthcare has resulted in significant advancements. One pressing concern in this field is diabetes, a disease that is rapidly increasing globally. According to the International Diabetes Federation (IDF), approximately 537 million adults aged 20 to 79 were living with diabetes in 2021, and this number is expected to rise to 643 million by 2030.

Despite growing awareness, many individuals remain undiagnosed due to limited access to medical services or a lack of understanding of the symptoms. Predictive analytics, powered by machine learning, provides a solution to this problem by offering affordable, scalable, and reliable tools that can be accessed by anyone with an internet connection.

This project aims to leverage data science and machine learning to develop a health prediction system that helps individuals take proactive steps toward managing their health.th.

1.2 Aim

The main goal of this project is to create a reliable, accurate, and user-friendly web application that predicts the likelihood of diabetes in patients based on medical parameters using machine learning models. This application is designed to assist in early diagnosis and encourage users to seek medical attention, ultimately contributing to improved management and treatment of diabetes.

1.3 Objectives

The objectives of this project are as follows:

- Gather a reliable dataset for diabetes prediction.
- Perform data preprocessing and cleaning to prepare the data for machine learning models.
- Analyze and visualize the data to gain valuable insights.
- Build and compare various machine learning models for diabetes prediction.
- Select the best-performing model based on performance metrics.
- Integrate the chosen model into a web application for public use.
- Provide a user-friendly interface that allows users to enter their health parameters and receive instant predictions.
- Ensure data security and privacy within the application.

2. Literature Survey

2.1 Background History

Diabetes has a long history, with the earliest known references dating back to 1500 BCE in Egyptian manuscripts. However, the scientific understanding and diagnosis of diabetes evolved in the 20th century. The discovery of insulin in 1921 revolutionized the treatment of diabetes. Over time, various diagnostic tests such as Fasting Blood Sugar (FBS), Oral Glucose Tolerance Test (OGTT), and Hemoglobin A1C were developed. With the advancement of computational techniques, researchers began exploring the use of artificial intelligence and machine learning in diagnosing and predicting diabetes. One of the most commonly used datasets in diabetes prediction research is the PIMA Indian Diabetes dataset, developed by the National Institute of Diabetes and Digestive and Kidney Diseases. This dataset includes various diagnostic measurements and the outcome indicating whether a patient has diabetes.

2.2 Related Work

Several studies have explored the use of machine learning for diabetes prediction. Some of the key works include:

- In "Diabetes Prediction Using Machine Learning" published in IJERT, researchers evaluated algorithms like Logistic Regression, Decision Tree, and Random Forest using the PIMA dataset.
- Another study titled "Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers" explored the use of ensemble learning to improve prediction accuracy.
- A paper published on SSRN emphasized the importance of preprocessing and feature selection in enhancing the model performance for diabetes prediction.
- Research also showed that combining multiple classifiers (e.g., Random Forest with SVM) can produce higher accuracy compared to single models.

These studies demonstrated the feasibility and effectiveness of machine learning techniques in predicting diabetes.

2.3 Limitations of Existing System

While existing research and systems have made significant progress, there are still several limitations:

- **Limited Accessibility**: Many existing models are not deployed as web applications and hence are not easily accessible to the general public.
- **Static Models**: Most models are static and do not learn from new data once deployed.
- **Data Imbalance**: Some models suffer from poor performance due to imbalanced datasets (i.e., fewer positive diabetes cases).
- **Lack of Interpretability**: Complex models like Random Forest and SVM lack transparency, making it difficult to explain predictions.
- **No Real-Time Feedback**: Many systems do not provide real-time or interactive prediction tools.

3. Idea/ Methodology

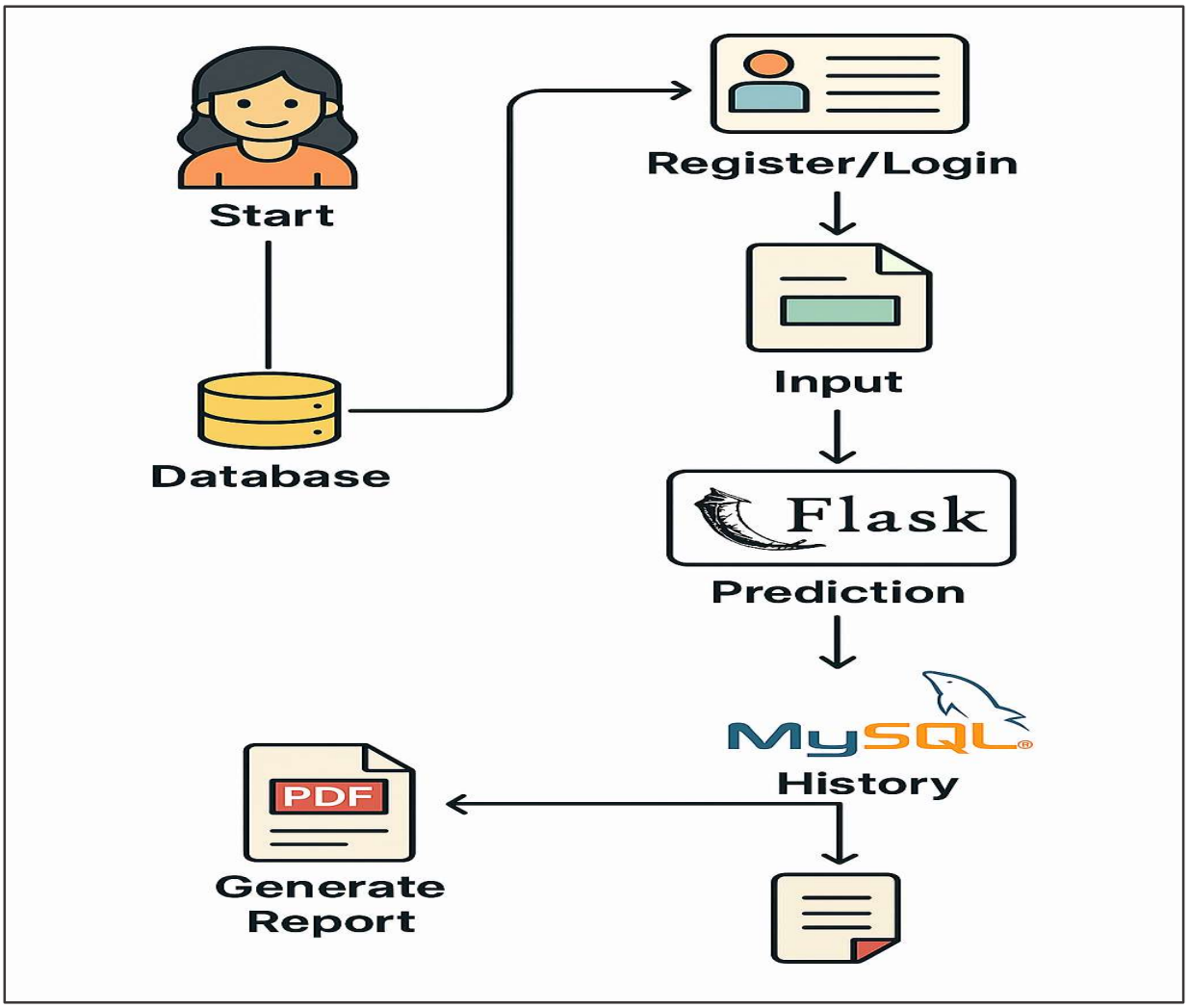
3.1 Basic Idea

The fundamental idea of this project is to use machine learning algorithms to analyze user-provided clinical data and predict whether a person is likely to have diabetes. The application provides a simple user interface where users input their health information such as age, glucose level, BMI, and more. The backend system processes this data using a trained machine learning model and returns a prediction.

This system is designed to be scalable and can be enhanced in the future to include more features such as live health monitoring through wearable devices.

3.2 Working

The project workflow includes the following steps:



1. Data Collection

   The dataset used in this project is the PIMA Indian Diabetes dataset, which includes 768 samples with 8 features:

- Pregnancies
- Glucose
- Blood Pressure
- Skin Thickness
- Insulin
- BMI
- Diabetes Pedigree Function
- Age
- Outcome (0 = non-diabetic, 1 = diabetic)

2. Data Preprocessing

- Handling Missing Values: Replacing zero values in certain features with the mean or median.
- Normalization: Feature scaling is done using MinMaxScaler or StandardScaler.
- Splitting: The data is divided into training and test sets (e.g., 80:20 ratio).

3. Model Building

   Various machine learning models are trained using the preprocessed data:

- Logistic Regression
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)
- Decision Tree
- Random Forest
- Naive Bayes

4. Model Evaluation

   Each model is evaluated using the following metrics:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

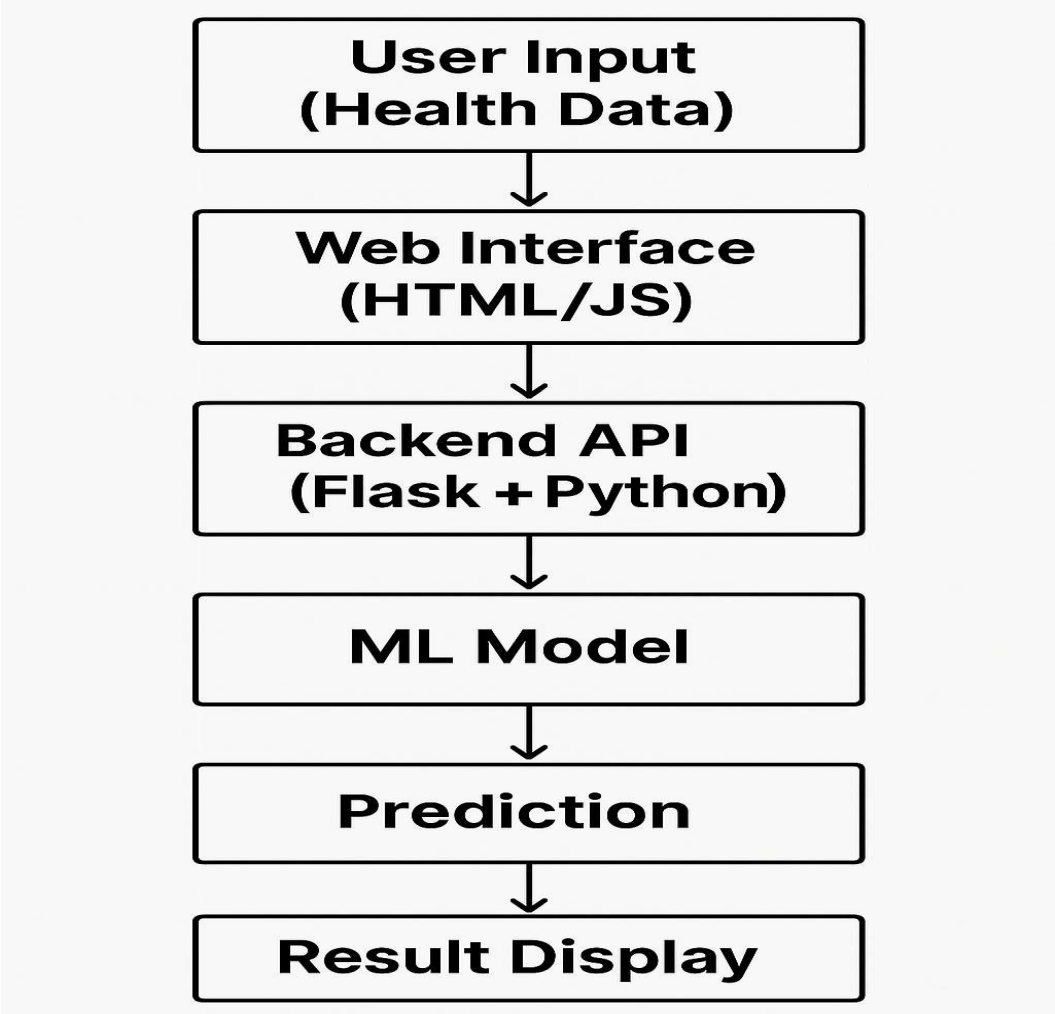
5. Model Selection

   The model with the highest overall performance is selected for deployment. In this case, Random Forest and SVM showed the best results.

6. Web Application Development

- **Frontend**: Developed using HTML, CSS, and JS.
- **Backend**: Developed using Python and Flask.
- **Integration**: The trained ML model is integrated into the Flask backend.
- **Deployment**: Hosted on a cloud platform

3.3 System Workflow



4. Results and Result Analysis

4.1 Introduction

The result analysis section evaluates the outcomes and effectiveness of the implemented machine learning-based diabetes prediction system. By examining key performance metrics, feature impact, model comparisons, and user interaction data, this section aims to assess the system's ability to assist in early detection, improve user accessibility, and support proactive healthcare management. The goal is to ensure that the solution not only delivers accurate predictions but also enhances the user experience and clinical applicability.

4.2 Model Performance and Accuracy

A critical factor in evaluating the success of the diabetes prediction system is its predictive accuracy and model reliability. Multiple machine learning models, including Logistic Regression, SVM, Random Forest, KNN, and Naive Bayes, were trained and tested on the PIMA Indian Diabetes dataset. Among these, Random Forest emerged as the best-performing model with an accuracy of 80%, followed by SVM (78%) and Logistic Regression (76%). The use of confusion matrices, ROC curves, and precision-recall scores further validated the model's performance, showcasing a strong ability to classify diabetic and non-diabetic patients accurately.

4.3 Feature Importance and Health Insights

The system also includes a feature importance analysis using Random Forest, which ranks health indicators based on their contribution to predictions. Features such as glucose level, BMI, age, and insulin levels were identified as the most influential. This insight helps healthcare practitioners and users understand the critical risk factors and focus on preventive measures accordingly. Such transparency increases the model's trustworthiness and educational value for end-users.

.4 User Experience and Accessibility

User satisfaction was a key goal during development. A web-based interface was designed to be simple, interactive, and responsive. Users—both medical professionals and general public—can input health data and receive real-time predictions. Feedback from test users highlighted the system's ease of use, time-saving benefits, and the convenience of receiving instant results without requiring lab-based diagnostics. This enhances accessibility, especially in rural or underserved regions where healthcare infrastructure may be limited.

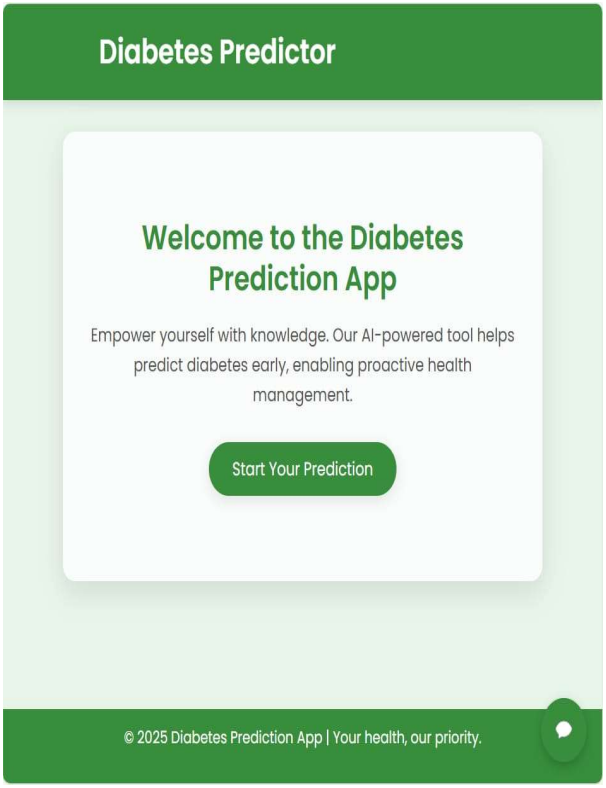
4.4 Early Detection and Preventive Health Support

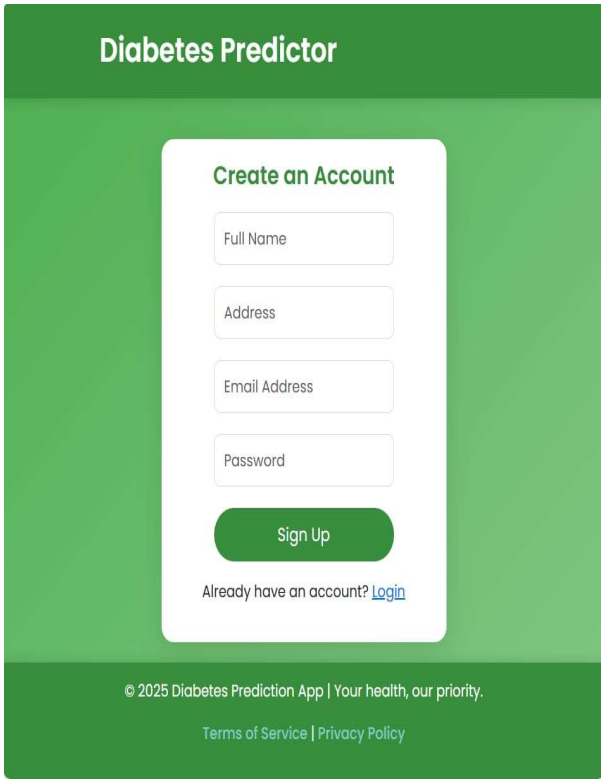
Early diagnosis is vital in managing chronic diseases like diabetes. The deployed system promotes early intervention by allowing individuals to perform preliminary screenings based on self-reported health data. By reducing reliance on traditional lab testing for initial risk assessment, the system contributes to faster identification of at-risk individuals, encouraging timely lifestyle changes or medical follow-up.

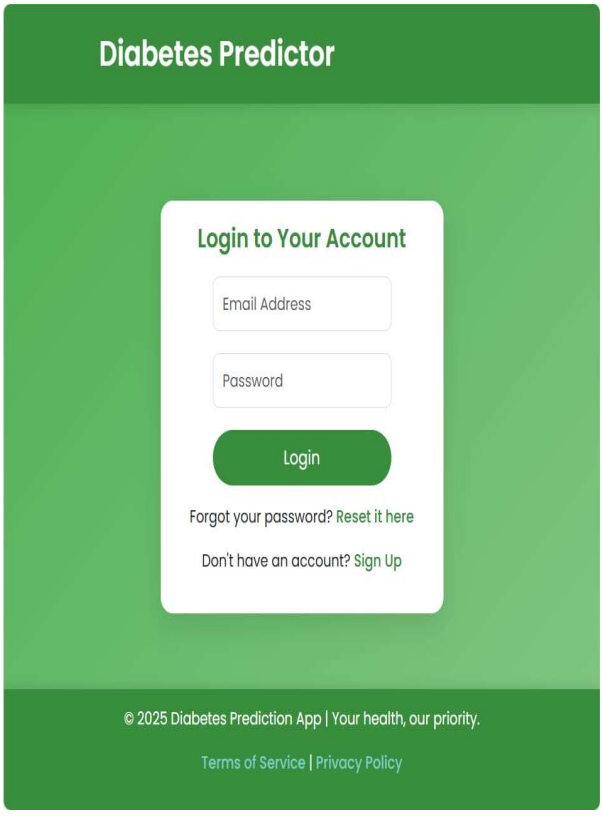
4.5 Scalability and Future Integration

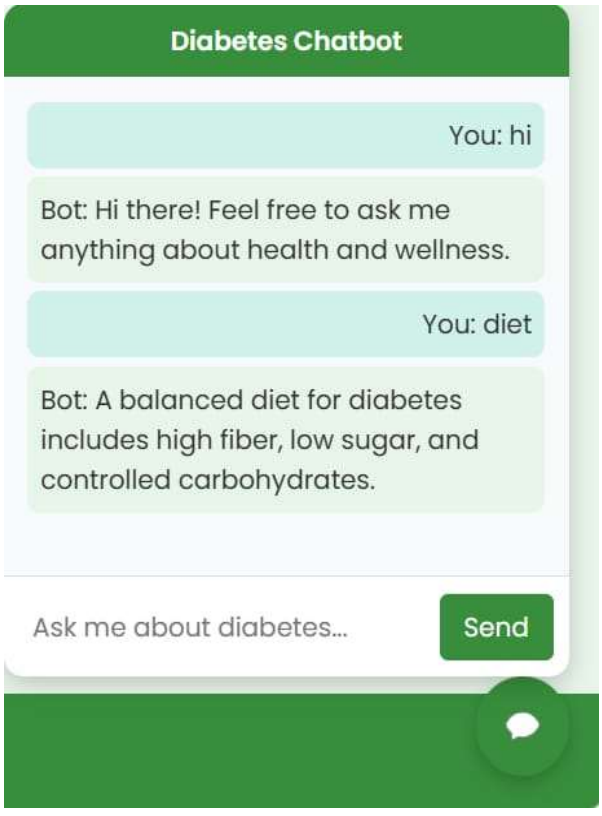
The diabetes prediction system is designed with scalability in mind. It can be expanded to include more datasets, patient demographics, and regional health statistics. Future integration with electronic health record (EHR) systems, mobile health apps, or wearable devices can further enhance real-time monitoring and predictive healthcare. Additionally, adding multilingual support and voice-enabled interfaces can help bridge digital literacy gaps in remote populations.

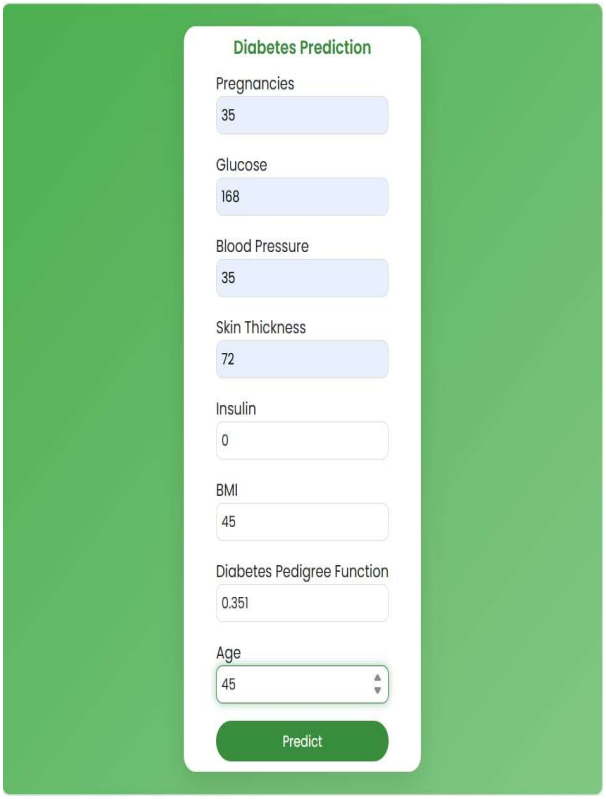
Screenshot

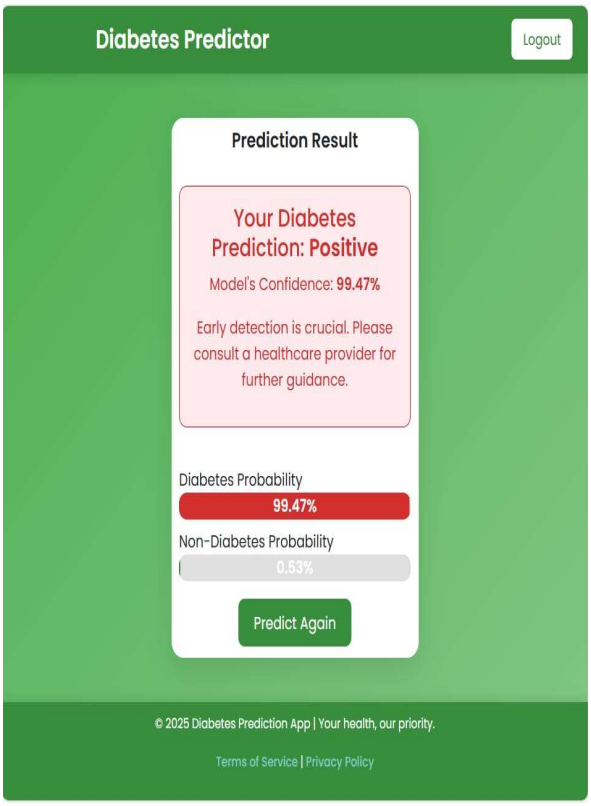

Diabetes Predictor
Welcome to the Diabetes Prediction App
Empower yourself with knowledge. Our AI-powered tool helps predict diabetes early, enabling proactive health management.
Start Your Prediction
© 2025 Diabetes Prediction App | Your health, our priority.

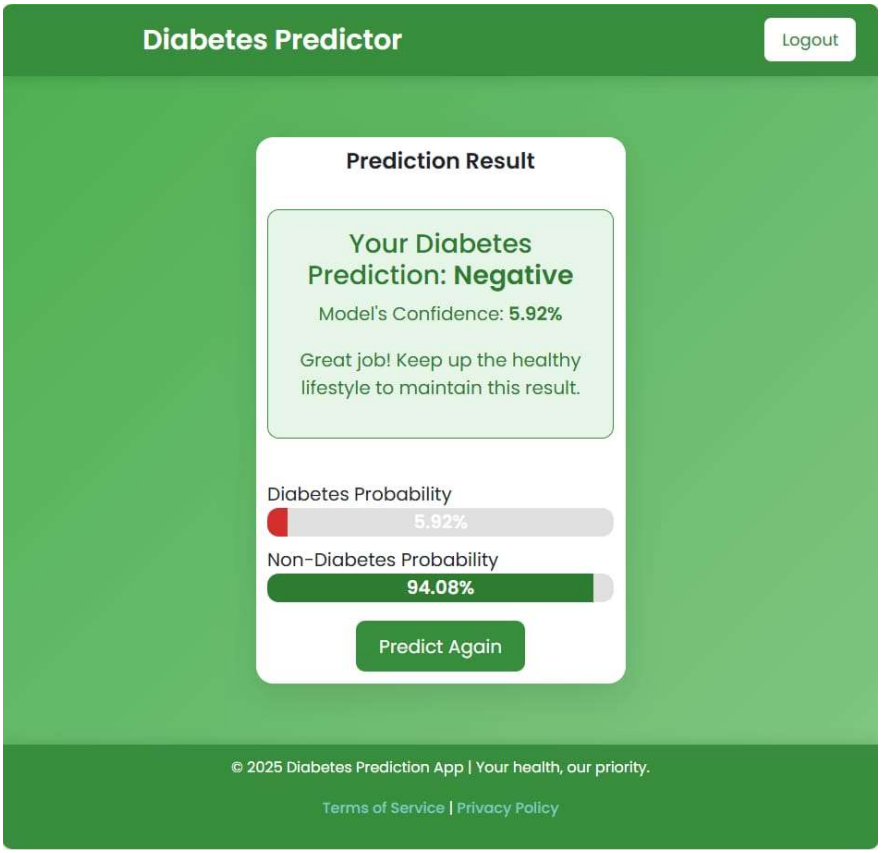

Diabetes Predictor
Create an Account
Full Name
Address
Email Address
Password
Sign Up
Already have an account? Login
© 2025 Diabetes Prediction App | Your health, our priority.
Terms of Service | Privacy Policy


Diabetes Predictor
Login to Your Account
Email Address
Password
Login
Forgot your password? Reset it here
Don't have an account? Sign Up
© 2025 Diabetes Prediction App | Your health, our priority.
Terms of Service | Privacy Policy


Diabetes Chatbot
You: hi
Bot: Hi there! Feel free to ask me anything about health and wellness.
You: diet
Bot: A balanced diet for diabetes includes high fiber, low sugar, and controlled carbohydrates.
Ask me about diabetes...
Send

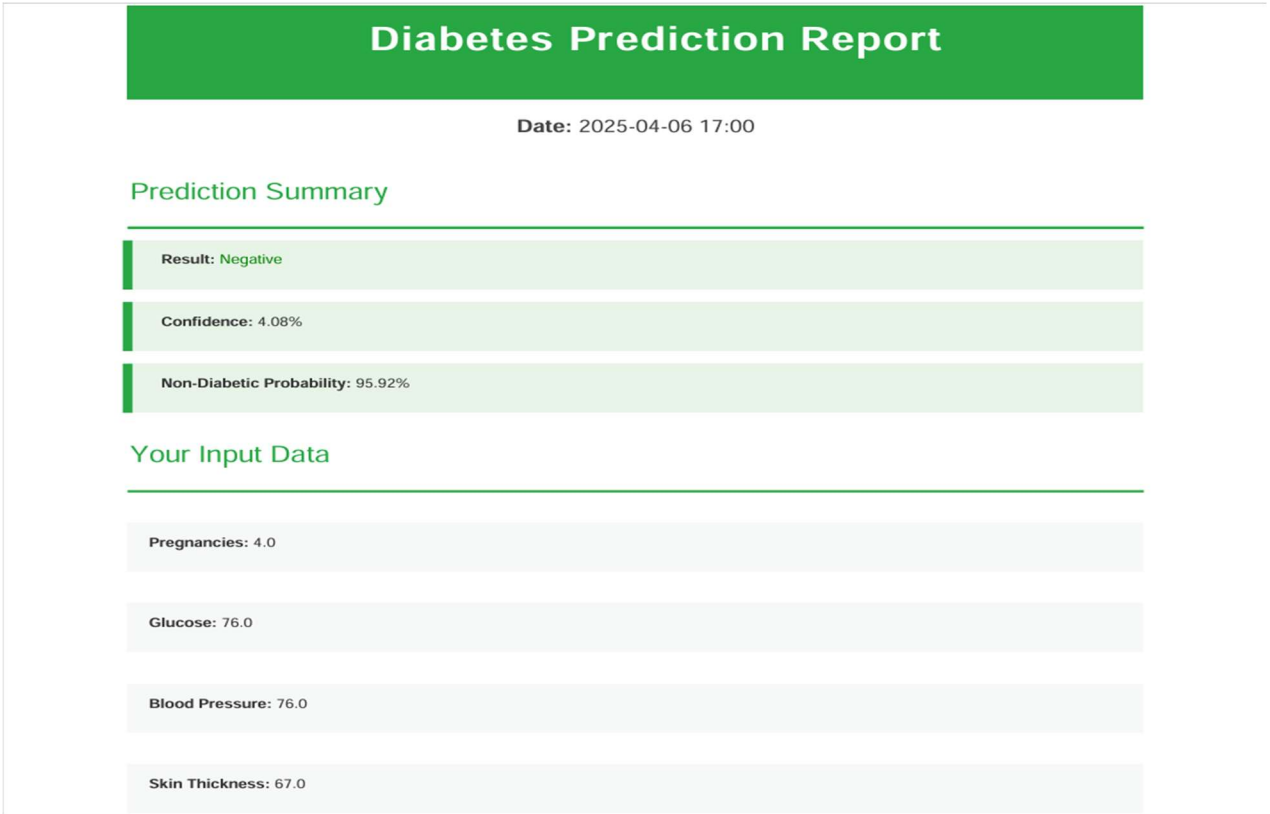

Diabetes Prediction
Pregnancies
35
Glucose
168
Blood Pressure
35
Skin Thickness
72
Insulin
0
BMI
45
Diabetes Pedigree Function
0.351
Age
45
Predict


Diabetes Predictor
Logout
Prediction Result
Your Diabetes Prediction: Positive
Model's Confidence: 99.47%
Early detection is crucial. Please consult a healthcare provider for further guidance.
Diabetes Probability
99.47%
Non-Diabetes Probability
0.53%
Predict Again
© 2025 Diabetes Prediction App | Your health, our priority.
Terms of Service | Privacy Policy


Diabetes Predictor
Logout
Prediction Result
Your Diabetes Prediction: Negative
Model's Confidence: 5.92%
Great job! Keep up the healthy lifestyle to maintain this result.
Diabetes Probability
5.92%
Non-Diabetes Probability
94.08%
Predict Again
© 2025 Diabetes Prediction App | Your health, our priority.
Terms of Service | Privacy Policy


Prediction History
All Time
Toggle Summary
Total Predictions: 84
Total Positive: 31
Total Negative: 53
Average Glucose: 52.8
Average BMI: 46.5
Date Glucose BMI Age Result Diabetic %
1 2025-04-06 85.0 28.1 56.0 Negative 12.27%
2 2025-04-06 67.0 65.0 7.0 Negative 35.14%
3 2025-04-06 67.0 65.0 7.0 Negative 35.14%
4 2025-04-06 67.0 65.0 7.0 Negative 35.14%
5 2025-04-06 67.0 65.0 7.0 Negative 35.14%
6 2025-04-06 85.0 28.1 56.0 Negative 12.27%

Report:


Diabetes Prediction Report
Date: 2025-04-06 17:00
Prediction Summary
Result: Negative
Confidence: 4.08%
Non-Diabetic Probability: 95.92%
Your Input Data
Pregnancies: 4.0
Glucose: 76.0
Blood Pressure: 76.0
Skin Thickness: 67.0

Insulin: 76.0

BMI: 9.0

DPF: 0.9

Age: 67.0

Health & Lifestyle Suggestions

- Adopt a balanced diet rich in fiber and low in sugar.
- Exercise regularly – even 30 mins of walking helps!
- Stay hydrated and manage your sleep schedule.
- Visit a healthcare provider for further evaluation.

5. Opportunity & Challenges

5.1 Opportunity

5.1.1 Healthcare Market Expansion

With the global rise in diabetes cases and the increasing focus on preventive healthcare, there is a significant market opportunity for expanding AI-driven diabetes prediction systems. As digital health adoption grows in both urban and rural areas, the system can be introduced in clinics, diagnostic centers, and wellness programs to support early diagnosis and risk prediction at scale.

5.1.2 Technological Advancements

The rapid evolution of Artificial Intelligence (AI), Machine Learning (ML), cloud computing, and wearable health tech presents new opportunities for improving system performance. Future upgrades can include real-time predictions, continuous health monitoring, and integration with Electronic Health Records (EHRs), enhancing clinical decision-making and personalized care.

5.1.3 Strategic Healthcare Collaborations

Collaborations with hospitals, telemedicine platforms, insurance companies, and public health departments offer opportunities to deploy the solution on a larger scale. Through partnerships, the system can reach diverse patient groups, gather real-world data, and support healthcare initiatives focused on chronic disease management and digital health transformation.

5.2 Challenges

5.2.1 Data Quality and Standardization

A major challenge is ensuring the input health data is accurate, consistent, and standardized. In real-world scenarios, users may input incorrect values, or health data may vary in format, making data preprocessing and validation critical to maintain model accuracy.

5.2.2 Privacy and Security Compliance

Handling sensitive medical information demands strict data protection measures, including compliance with HIPAA, GDPR, or national data privacy regulations. Implementing robust data encryption, user consent mechanisms, and secure cloud infrastructure is vital to build and maintain trust among users and stakeholder

6. Conclusion & Future Scope

6.1 Conclusion :

This project highlights the practical application of machine learning in building an intelligent, web-based Diabetes Prediction System aimed at early detection and awareness. Using the PIMA Indian Diabetes Dataset, multiple classification algorithms were trained and evaluated, with the Random Forest model achieving the highest accuracy at 80%. The system offers a user-friendly interface that enables individuals to input health parameters and receive instant predictions, empowering users to make informed health decisions.
The application stands as a cost-effective, accessible, and scalable solution for preliminary diabetes screening, particularly useful in remote or underserved regions. By leveraging data-driven insights, it contributes meaningfully to preventive healthcare and supports the goal of early diagnosis.

6.2 Future Scope :

To further enhance the functionality, usability, and impact of the system, several future developments are proposed:

- Cloud Deployment: Hosting the application on platforms like AWS or Google Cloud to improve scalability, availability, and real-time performance.
- Wearable Device Integration: Connecting with fitness trackers and smartwatches to collect real-time physiological data for continuous monitoring and improved prediction accuracy.
- Multi-Disease Detection: Extending the platform's capabilities to include other chronic diseases such as hypertension, cardiovascular disorders, and obesity.
- AI Chatbot Support: Implementing intelligent chatbots for guidance, feedback, and basic health counseling, making the system more interactive and supportive.
- Continuous Learning: Incorporating mechanisms for real-time data ingestion and model retraining, allowing the system to stay updated with changing population health trends.
- Privacy & Security Enhancements: Ensuring the system complies with medical data privacy standards to build trust and user confidence.

7. References

7.1 References links:

[1] UCI Machine Learning Repository, "Pima Indians Diabetes Database," *Kaggle*, 2020. [Online]. Available: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

[2] International Diabetes Federation, *IDF Diabetes Atlas*, 10th ed., 2021. [Online]. Available: https://diabetesatlas.org/

[3] J. Brownlee, "A Tour of Machine Learning Algorithms," *Machine Learning Mastery*, 2020. [Online]. Available: https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

[4] World Health Organization, "Diabetes," [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes

[5] R. T. Prasad, S. B. Nagajyothi, and G. Harika, "Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 06, no. 06, pp. 1–5, Jun. 2022.

[6] A. Kadam and S. Zade, "Diabetes Prediction Using Machine Learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 08, no. 06, pp. 3526–3530, Jun. 2021.

[7] D. Shinde, P. Patil, A. Varpe, P. Ghadge, and R. Gawade, "Diabetes Prediction using Machine Learning," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 12, pp. 1–4, Dec. 2021.

[8] E. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019. doi: 10.1038/s41591-018-0300-7

[9] A. Smith et al., "Application of Machine Learning Algorithms in Healthcare: Diabetes Prediction Model," *Journal of Medical Systems*, vol. 46, no. 1, 2022. doi: 10.1007/s10916-021-01777-4

[10] M. R. Patel and A. P. Zaveri, "Diabetes Prediction Using Machine Learning," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 11, pp. 325–330, 2020.

[11] P. S. Kumar and D. D. Sharma, "Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers," *SSRN Electronic Journal*, 2021. doi: 10.2139/ssrn.3763744

[12] R. Smith et al., "A Review on Machine Learning Algorithms for Diabetes Prediction," *International Journal of Computer Applications*, vol. 180, no. 38, pp. 1–5, 2018.