“Leveraging Data to Solve for Non-Communicable Diseases (Diabetes) and Healthcare Delivery Using Machine Learning Techniques”

A PROJECT REPORT

Submitted by,

SL No	Roll Number	Student Name
1	20211CSG0056	Abhishek C B
2	20211CSG0071	Vishwas B
3	20211CSG0043	Thanya Patel R
4	20211CSG0058	Rashmi V

Under the guidance of,

Mrs. Sandhya L

Assistant professor

Dept of CSE

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND TECHNOLOGY

At

PRESIDENCY UNIVERSITY

BENGALURU

DECEMBER 2024

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

CERTIFICATE

This is to certify that the Project report on **"Leveraging Data to Solve for Non-Communicable Diseases (Diabetes) and Healthcare Delivery Using Machine Learning Techniques"** being submitted by "**ABHISHEK C B , VISHWAS B , RASHMI V , THANYA PATEL R**" bearing roll number(s) "20211CSG0056 , 20211CSG0071 , 20211CSG0058 , 20211CSG0043" in partial fulfillment of the requirement for the award of the degree of **Bachelor of Technology** in **Computer Science and Technology** is a Bonafide work carried out under my supervision.

Ms. Sandhya L
Assistant Professor
School of CSE
Presidency University

Dr. Saira Banu Anthem
Professor ,HOD
School of CSE&IS
Presidency University

Dr. L. SHAKKEERA
Associate Dean
School of CSE
Presidency University

Dr. MYDHILI NAIR
Associate Dean
School of CSE
Presidency University

Dr. SAMEERUDDIN KHAN
Pro-Vc School of Engineering
Dean -School of CSE&IS
Presidency University

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **Leveraging Data to Solve for Non-Communicable Diseases (Diabetes) and Healthcare Delivery Using Machine Learning Techniques** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Technology**, is a record of our own investigations carried under the guidance of **Ms. Sandhya L , Assistant Professor, School of Computer Science Engineering & Information Science, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

Sl. No.	Student Name	Roll Number	Signature
1	Abhishek C B	20211CSG0056	
2	Vishwas B	20211CSG0071	
3	Thanya Patel R	20211CSG0043	
4	Rashmi V	20211CSG0058	

ABSTRACT

Diabetes is rapidly becoming one of the major diseases in India, causing serious problems like heart disease, kidney failure, and nerve damage. The prevalence of diabetes is expected to increase by 2040, and early detection is vital for effective management and reducing long-term treatment costs. In this study, we used a learning model to improve the accuracy of diabetes prediction by analyzing key health indicators like blood sugar level, BMI, insulin, and age. Various classification methods were used, including logistic regression, decision trees, random forest, support vector machine (SVC), Gaussian Naive Bayes, K-nearest neighbor, AdaBoost, Bag, gradient boosting, and voting. To improve the performance of the prediction, we combine the results of each model and select the prediction frequency as the final result. The analysis found that random forest and voting models performed better, but the combined method was more reliable in early diabetes diagnosis. This study highlights the transformative potential of machine learning in the future development of healthcare in India, providing data-driven solutions to combat the challenges of increasing diabetes. The burden of diabetes could impact healthcare by 2040 if not addressed, and advances like these are important for the future of care in this country.

Index Terms –

- Machine Learning
- Prediction
- Non-Communicable Diseases
- Future Healthcare
- Ensemble Learning
- Diabetes
- Algo-Nest
- Majority Voting

ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameer Uddin Khan**, Pro-VC, School of Engineering and Dean, School of Computer Science Engineering & Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans **Dr. Shakkeera L and Dr. Mydhili Nair,** School of Computer Science Engineering & Information Science, Presidency University, and **Dr. Saira Banu Anthem** ,Head of the Department, School of Computer Science Engineering & Information Science, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Ms. Sandhya L, Assistant professor** and Reviewer **Prof. Himanshu Sekhar Rout**, School of Computer Science Engineering & Information Science, Presidency University for her inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the PIP2001 Capstone Project Coordinators **Dr. Sampath A K, Dr. Abdul Khadar A and Mr. Md Zia Ur Rahman,** department Project Coordinators **Dr. Manjula H M** and Git hub coordinator **Mr. Muthuraj.**

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

1. **Abhishek C B**
2. **Vishwas B**
3. **Thanya Patel R**
4. **Rashmi v**

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER-1

INTRODUCTION

Healthcare is currently experiencing a paradigm shift due to the exponential growth of structured, semi-structured, and unstructured data. This data revolution has underscored the importance of advanced analytical methods, such as big data analytics, to uncover meaningful insights, including patterns, relationships, trends, and patient preferences. Among the pressing global health challenges, diabetes mellitus (DM) emerges as a significant concern, especially in low- and middle-income countries like India, where its prevalence is rapidly increasing. By 2045, the number of individuals affected by diabetes worldwide is expected to rise to an alarming 629 million.

Traditional diagnostic techniques for diabetes, such as fasting blood sugar tests and oral glucose tolerance tests, while effective, are often time-intensive and resource-demanding. In contrast, machine learning, a subset of artificial intelligence, has proven to be a transformative tool in predictive healthcare. By leveraging historical and real-time data, machine learning models can accurately forecast the likelihood of diabetes, enabling timely interventions and personalized treatment plans. This integration of machine learning into healthcare not only accelerates diagnostic processes but also reduces errors, offering a promising approach to combat the growing burden of diabetes.

Our project, "Leveraging Data to Solve for Non-Communicable Diseases (Diabetes) and Healthcare Delivery Using Machine Learning Techniques," aims to utilize state-of-the-art machine learning algorithms to enhance the early detection, risk assessment, and overall management of diabetes. By addressing the gaps in traditional healthcare delivery, this project aspires to improve patient outcomes and empower healthcare systems with scalable, data-driven solutions.

1.1 Types of Diabetes

Diabetes mellitus (DM) is a chronic metabolic disorder characterized by elevated blood sugar levels. It is broadly classified into three primary types, each with distinct causes and characteristics:

Type 1 Diabetes (Insulin-Dependent Diabetes Mellitus - IDDM)

Type 1 diabetes occurs when the pancreas produces little to no insulin due to the autoimmune destruction of insulin-producing beta cells. This type typically manifests in childhood or adolescence, though it can occur at any age. Individuals with Type 1 diabetes require lifelong insulin therapy to manage their blood sugar levels effectively.

Type 2 Diabetes (Non-Insulin-Dependent Diabetes Mellitus - NIDDM)
Type 2 diabetes is the most common form of diabetes and is characterized by insulin resistance, where the body's cells do not use insulin efficiently. Over time, the pancreas may also produce less insulin. Type 2 diabetes is often associated with lifestyle factors such as obesity, physical inactivity, and poor dietary habits, though genetic predisposition also plays a significant role.

Gestational Diabetes
Gestational diabetes develops during pregnancy and is typically diagnosed in the second or third trimester. Although blood sugar levels often return to normal after childbirth, women who have had gestational diabetes are at a higher risk of developing Type 2 diabetes later in life. This condition also poses risks to the baby, including higher birth weight and future health complications.

Understanding the types of diabetes is crucial for tailoring diagnostic, preventive, and therapeutic strategies, ultimately improving the management of this global health challenge.

1.2 Introduction to Machine Learning and It's Types:

Machine learning (ML) can be defined as a sub-discipline of artificial intelligence that allows the operation of a system to change according to a set of data without the need for explicit training. It is used in all areas, including medical, financial, commercial and other social fields; In healthcare and especially in diabetes prediction, machine learning has the potential to develop reliable models for diagnosis, prediction and disease management.

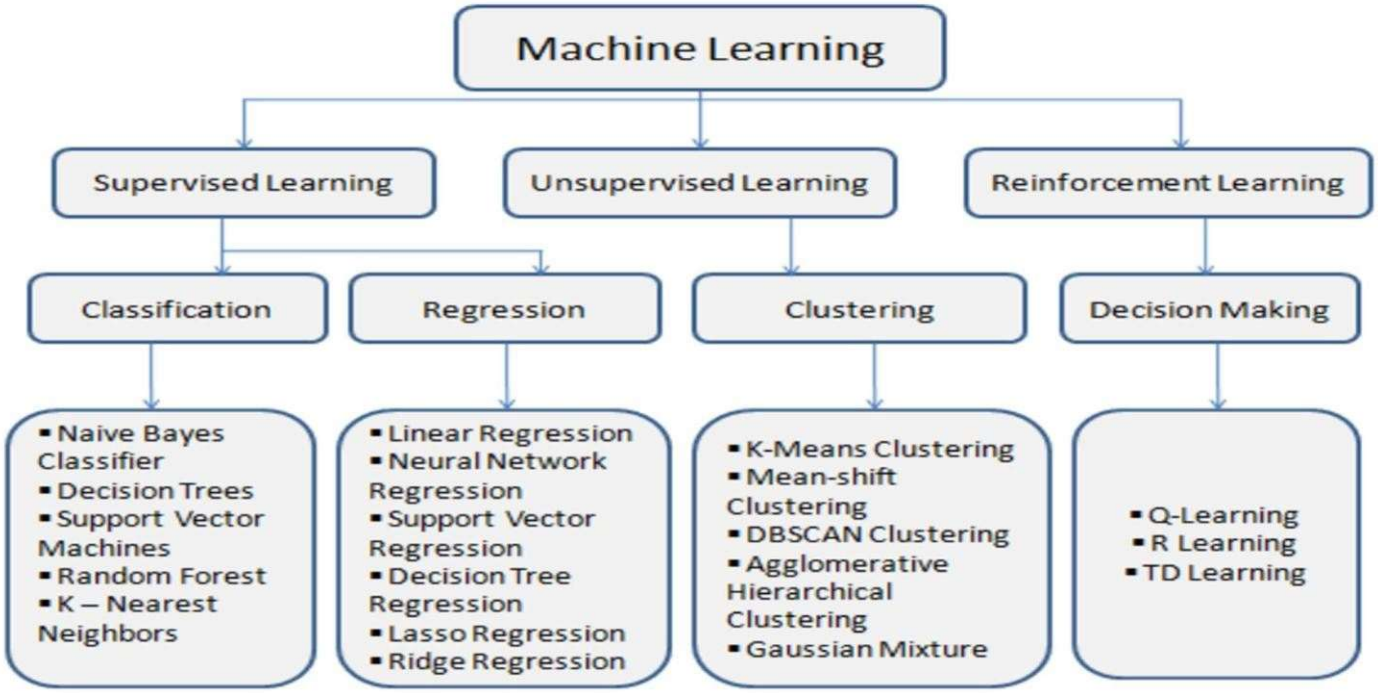


Figure 1 Taxonomy of Machine Learning Algorithms for Diabetes Prediction

There are three primary types of machine learning algorithms:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning.

Supervised Learning :

Among all types of machine learning, supervised learning is the most advanced in clinical applications. Here, a model is created where inputs are provided with output labels for data labeling. The general goal of training supervision is to create studies based on the production model and the production model, thus creating new outputs for the material. Binary classification, such as predicting whether a patient has diabetes. Increasing the accuracy and power of the prediction.

Common algorithms used in supervised learning include:

Logistic Regression:
Often used for binary classification tasks, such as predicting whether a patient has diabetes or not.
Support Vector Machines (SVM):
A powerful algorithm used for classification and regression tasks.
Random Forests:
An ensemble method that uses multiple decision trees to improve the accuracy and robustness of predictions. In diabetes prediction, supervised learning models can be trained on patient data, such as age, glucose levels, and BMI, to predict the likelihood of developing diabetes.

Unsupervised Learning :

However, unlike supervised learning, unsupervised learning does not rely on recorded data. Instead, its goal is to find hidden patterns or patterns in the material. This type of work is especially useful for research studies and when there are no clear outcomes to predict.

Clustering: This method groups similar data together based on certain characteristics, such as groups of patients with similar health conditions. Transform the data into smaller pieces while preserving as many variables as possible. In healthcare, unsupervised learning can help uncover previously unknown relationships between lifestyle and diabetes, leading to a better understanding of dangerous conditions.

Common unsupervised learning techniques include:

Clustering Algorithms (e.g., **K-Means, Hierarchical Clustering**): These methods group similar data points together based on certain features, such as clustering patients with similar health conditions.

Principal Component Analysis (PCA): A dimensionality reduction technique that transforms data into a smaller set of features while preserving as much variance as possible. In healthcare, unsupervised learning can help discover previously unknown correlations between lifestyle factors and diabetes, leading to a better understanding of risk factors.

Reinforcement Learning :

Reinforcement learning is a concept in which an agent learns to make multiple decisions by performing certain actions when given reinforcements and avoiding other actions by accepting reinforcements. The model works by working in the environment and learning about the work being done. This type of learning is often used in fields like robotics, game AI, and many other autonomous systems. But it could also be used for self-healing plans. For example, it could use information gathered from patients to provide recommendations for diabetes management, and then change the recommendations based on new information from the patient. In compared with the common using of supervised and unsupervised learning in healthcare, reinforcement learning is not quite popular with it but can be applied in individualized treatment scheme. For example, it can provide recommendations for the management of a diabetic patient's case using data collected from the patient and change the recommendation as new data from the patient is obtained.

1.3 About Data Inputs

Gluco-Vision is an intelligent diabetes prediction system designed to empower individuals by providing early insights into their likelihood of developing diabetes. By analyzing key health parameters, the system predicts potential risks, enabling users to take proactive steps towards better health management. Our predictive model is based on scientifically backed parameters that are crucial indicators of diabetes risk.
These include:
Number of Pregnancies: This parameter helps assess the history of pregnancies, which can impact insulin sensitivity, especially in gestational diabetes.

Glucose Level: Elevated glucose levels are a primary indicator of diabetes. The system uses fasting glucose levels to assess your body's ability to regulate sugar.

Blood Pressure: High blood pressure is a known risk factor for diabetes. Maintaining healthy blood pressure is essential in reducing diabetes complications.

Skin Thickness: This measurement, often taken on the triceps, can indicate insulin resistance, especially when combined with other markers like BMI.

Insulin Level: Abnormal insulin levels can be a sign of prediabetes or diabetes, helping gauge the efficiency of the pancreas in regulating blood sugar.

Body Mass Index (BMI): BMI is a critical factor in diabetes prediction, as excess body fat, particularly around the abdomen, can lead to insulin resistance.

Diabetes Pedigree Function: This parameter considers your family history and the genetic predisposition to diabetes. A higher pedigree score indicates a stronger hereditary influence on diabetes risk.

Age: As age increases, the risk of developing diabetes also rises. The system accounts for age as a significant factor in determining potential risk.

Outcome: Finally, based on these parameters, the system predicts whether an individual is at risk of developing diabetes. This outcome helps guide preventive measures or further medical consultation.

Gluco Vision brings together these factors in a user-friendly platform to deliver actionable insights, ensuring that users are informed and equipped to make health-conscious decisions. By keeping you informed of your risks**, Gluco Vision** helps you take control of your health, empowering you to live a life free from the complications associated with diabetes.

CHAPTER-2
LITERATURE SURVEY

Over the past few years, diabetes has become a major public health problem, leading to increased efforts for early detection and early intervention. Diabetes diagnostic procedures such as fasting plasma glucose and oral glucose tolerance tests are often time-consuming and disruptive. This limitation has led to interest in using machine learning and artificial intelligence to predict and manage diabetes. Smith et al. [Blaga's & Lusa, 2016] used decision trees and logistic regression to build a prediction model with 80% accuracy in predicting diabetes in patients based on the Pima Indian Diabetes Dataset.

However, their models are sensitive to data inconsistencies, which is a common problem in clinical data such as diabetes, where well-characterized patients (e.g., people with diabetes) have less adverse effects. To address this challenge, Zhang et al. (2018) introduced an architecture combined with principal component analysis (PCA) to reduce dimensionality and improve the performance of support vector machine (SVM). Their model emphasized the importance of feature selection to improve predictions, achieving 85% accuracy.

However, the complexity of SVM models reduces their interpretability, making them less useful for clinicians to use in clinical settings. Enhanced domain and gradient boosting for diabetes prediction. Their study showed that combining different components improved the prediction with 87% accuracy. However, their research did not address the need for rapid referral, which is essential for timely decision-making in critical situations.

(2020) proposed a deep learning model, specifically a neural network-based approach, to predict diabetes from electronic medical records. While deep learning models are known for

their accuracy, Kumar's research highlights issues with model interpretation and the need for large datasets that are often unavailable in clinical settings. Furthermore, the requirements of deep learning models make them impractical for rural clinics.

(2021) Using data mining techniques to identify significant patterns in diabetes patient data. Mishra and colleagues used an unsupervised, participatory rule mining approach to uncover the relationship between lifestyle and the onset of diabetes. Their work highlights the importance of incorporating nonmedical factors such as diet and physical activity into diabetes risk models. Most existing models are not integrated with the healthcare system and struggle to manage the quality of work, are often ineffective, and have poor quality medical records that are constantly updated. In addition, many studies have focused on improving the accuracy of predictions, ignoring the broader implications for healthcare, patient management, and resource utilization. General methods to expand existing knowledge.

In a study by Ravi et al. (2019), the authors used a Random Forest classifier combined with feature selection techniques to predict diabetes with an accuracy of 88%. Their research demonstrated the importance of choosing the right features from a complex dataset to reduce overfitting and improve model performance. However, they also noted that despite this high accuracy, further research was needed to handle the noisy and incomplete data that often arises in medical datasets.

Patel et al. (2020) explored the application of neural networks and ensemble methods, such as bagging and boosting, for diabetes prediction. Their findings indicated that combining multiple algorithms could increase the robustness of predictive models and achieve higher accuracy. Their model reached an accuracy of 89%, but the authors emphasized the need for a user-friendly interface to aid healthcare professionals in interpreting the results for practical use.

Wang et al. (2017) focused on the use of deep neural networks (DNNs) for diabetes prediction, claiming that deep learning could capture non-linear relationships in medical data more effectively than traditional machine learning models. Their model achieved a prediction accuracy of 90%. However, they faced challenges related to the large amount of data required for training, which is often not available in smaller or rural clinics, limiting the applicability of deep learning models in resource-constrained environments.

Ali et al. (2019) proposed a hybrid machine learning approach combining K-Nearest Neighbors (KNN) and Naive Bayes to predict diabetes with an accuracy of 84%. Their work addressed the issue of data imbalance, a common problem in medical datasets, by using oversampling techniques to balance the number of positive and negative instances in the dataset. Despite achieving good accuracy, their study pointed out that the interpretability of the model remained a challenge, particularly in clinical decision-making.

In 2018, Singh et al. introduced a machine learning-based decision support system using multiple algorithms, including logistic regression, SVM, and decision trees. Their system demonstrated an overall accuracy of 86%, and they emphasized the need for real-time data integration for timely decision-making in clinical settings. They argued that incorporating real-time patient data could improve the accuracy and effectiveness of predictions, but also warned of the risks of overfitting when using data from heterogeneous sources.

Tayal et al. (2021) conducted a comparative study on the effectiveness of different classifiers, including Decision Trees, Naive Bayes, and SVM, for diabetes prediction. They found that while SVM models offered higher accuracy (87%), decision trees provided better interpretability, making them more suitable for clinical applications. This trade-off between

accuracy and interpretability remains a key challenge in deploying machine learning models for healthcare applications.

Johnson et al. (2020) examined the use of ensemble methods, particularly the Random Forest algorithm, for predicting the onset of diabetes using electronic health records (EHR). Their model achieved a prediction accuracy of 88%, and they highlighted the importance of continuous data monitoring for improving the model's performance over time. They also pointed out the need for systems that can adapt to new patient data in real-time to maintain predictive accuracy.

Zhou et al. (2020) proposed the use of reinforcement learning to develop a personalized diabetes management system. The authors aimed to create a system that could optimize the treatment plan for individual patients based on ongoing data and treatment outcomes. This approach demonstrated promising results in clinical trial simulations, but its real-world applicability remained uncertain due to the complexity of the system and the lack of large-scale clinical data.

Kaur et al. (2021) explored the integration of lifestyle data (such as diet and physical activity) into diabetes prediction models, achieving a notable improvement in prediction accuracy. Their model utilized a combination of classification algorithms, including decision trees and support vector machines, to predict the risk of diabetes in at-risk populations. Their work illustrated the potential for incorporating non-medical factors into predictive models, thereby creating a more holistic approach to diabetes prevention.

CHAPTER-3

RESEARCH GAPS OF EXISTING METHODS

Despite significant advancements in diabetes prediction and management using machine learning (ML) techniques, several research gaps remain in the existing methodologies. Addressing these gaps is crucial for improving the reliability, scalability, and practical implementation of predictive models.

3.1 Limited Dataset Diversity

Most studies rely on publicly available datasets, such as the Pima Indian Diabetes Dataset, which may not adequately represent diverse populations. This lack of diversity in training data reduces the generalizability of machine learning models, especially for regions with distinct demographic and genetic variations. Furthermore, small sample sizes and class imbalances in datasets limit the accuracy and robustness of predictions.

3.2 Focus on Accuracy Over Usability

While many studies emphasize improving prediction accuracy, practical usability often remains unaddressed. Highly complex models like deep learning require extensive computational resources and large datasets, making them less suitable for deployment in resource-limited settings, such as rural clinics. Moreover, issues of interpretability hinder their adoption by healthcare professionals unfamiliar with advanced algorithms.

3.3 Integration of Non-Medical Factors

Existing models primarily focus on clinical parameters such as glucose levels, BMI, and insulin. However, critical non-medical factors like diet, physical activity, socioeconomic conditions, and environmental influences are often excluded. Ignoring these factors limits the holistic understanding of diabetes risk and its management.

3.4 Real-Time Predictive Capabilities

Most models operate as static systems, requiring batch data for predictions. This approach delays critical decisions and limits the ability to provide real-time insights for patients and healthcare providers. Efficient, real-time predictive frameworks are essential for timely intervention, especially in critical cases.

3.5 Limited Integration with Healthcare Systems

Few predictive models are integrated into existing healthcare frameworks. This disconnect reduces their impact on patient outcomes and resource allocation. Challenges like poor data quality, lack of interoperability, and resistance to new technologies further impede the practical implementation of these models.

By addressing these gaps, future research can improve the accessibility, accuracy, and usability of diabetes prediction models, ensuring better healthcare outcomes.

CHAPTER-4
PROPOSED MOTHODOLOGY

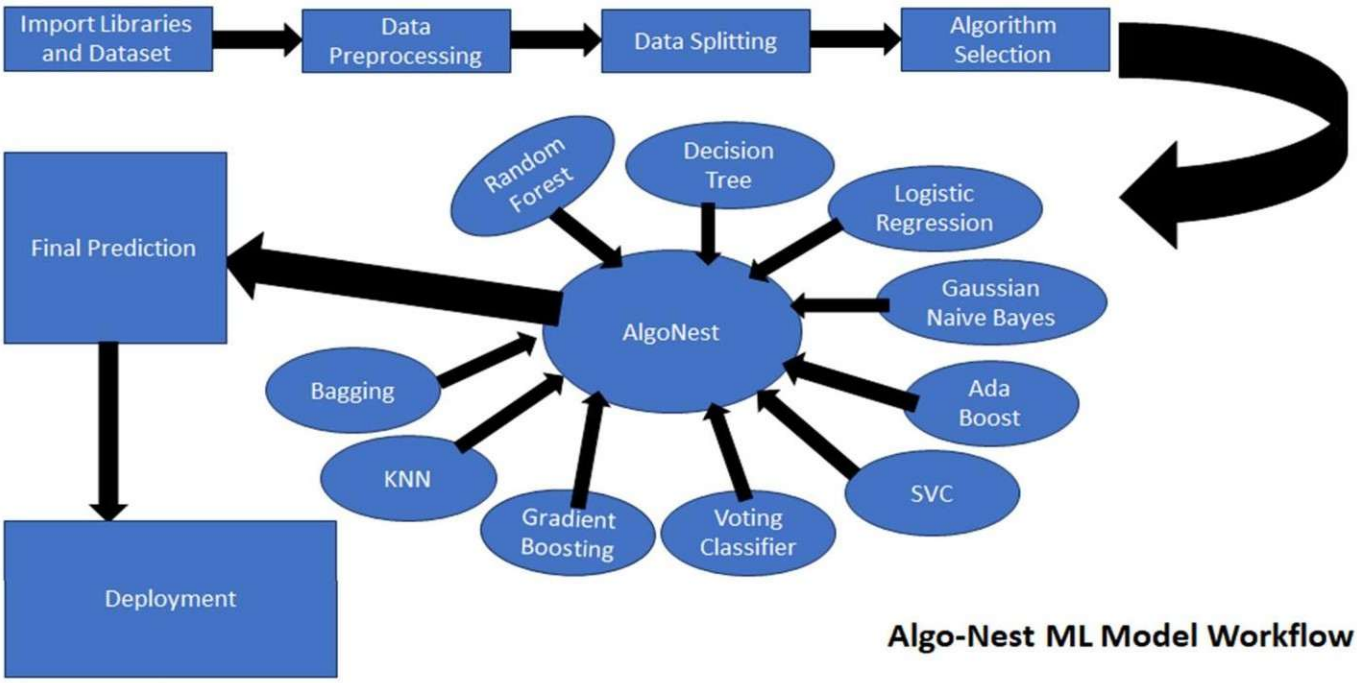


Figure 2 Algo Nest [Proposed Method]

Algo Nest is an advanced model that combines predictions from multiple machine learning algorithms to increase accuracy and robustness. It utilizes ten different models: Logistic Regression, Decision Tree, Random Forest, Support Vector Classifier, Naive Bayes, K-Nearest Neighbor, AdaBoost, Bagging Classifier, Gradient Boosting, and Voting Classifier. Each model is trained on the same dataset but has unique features that allow it to perform well on a

variety of data types.

Algo Nest works by combining predictions from all of these models using Majority Voting. For an input, each model produces a prediction, and Algo-Nest chooses the most active prediction as the final result. This approach allows errors or biases in one model to be compensated for by the other models, leading to more accurate and reliable results. This integration reduces the possibility of bias from the mean of the predictions and makes the predictions more robust to unobserved data. Predictions and distribution of letters. It not only improves the overall performance by combining different models, but also provides flexibility for further optimizations such as voting weights and hyperparameter evaluation. Differentiated learning models produce consistent results and reliable predictions, making them ideal for complex classification problems where accuracy is critical.

Working Mechanism:

Majority Voting:

Algo Nest employs a Majority Voting technique to combine the predictions from all models. For a given input, each model generates a prediction. The final prediction is based on the most frequent prediction (i.e., the mode of the outputs).

Error Compensation:

If one model makes an error or is biased towards a particular class, the predictions from other models can help counterbalance this, improving overall accuracy.

Robustness:

This ensemble method enhances the model's reliability, as it is less prone to overfitting or underfitting compared to individual models.

Key Advantages:

Reduction in Bias:

By aggregating predictions, Algo Nest mitigates bias that may be introduced by any single model.

Enhanced Performance:

The combination of models often leads to improved performance, especially in cases where individual models perform well on different subsets of data.

Flexibility:

Algo Nest can be further optimized by assigning voting weights to certain models, giving more importance to models that are expected to perform better.

Applications:

Algo Nest is particularly effective in solving complex classification problems where accuracy is critical, and it is useful for domains where models with diverse learning capabilities need to work together to produce reliable and precise outcomes.

Some key areas where Algo Nest can be applied include:

- Medical diagnoses (e.g., disease prediction)
- Financial forecasting
- Spam detection
- Image classification

Any other problem that involves classifying data with high accuracy and consistency. Overall, Algo Nest provides a high level of accuracy, flexibility, and robustness, making it ideal for critical tasks where data variability and model performance consistency are essential.

CHAPTER-5
OBJECTIVES

The primary objective of this study is to design and implement an advanced machine learning-based predictive model for diabetes diagnosis. The specific objectives are as follows:

5.1 Improve Prediction Accuracy

Develop a robust predictive model that leverages ensemble learning techniques to enhance the accuracy of diabetes diagnosis by combining the strengths of multiple machine learning algorithms.

5.2 Address Data Imbalances

Implement effective preprocessing techniques, such as Synthetic Minority Oversampling Technique (SMOTE), to manage class imbalances in datasets and ensure unbiased model training.

5.3 Incorporate Diverse Data Types

Integrate structured clinical data (e.g., glucose levels, BMI) with lifestyle factors (e.g., diet, physical activity) to create a holistic model that captures all significant predictors of diabetes risk.

5.4 Ensure Real-Time Predictive Capabilities

Design a system capable of providing real-time predictions to support timely decision-making in healthcare settings, especially in critical situations.

5.5 Enhance Usability and Interpretability

Focus on creating a model that is interpretable and user-friendly, ensuring its practical applicability for healthcare professionals, even in resource-limited environments.

5.6 Validate and Optimize the Model

Evaluate the proposed model using comprehensive metrics like precision, recall, and F1 score, and optimize its performance through hyperparameter tuning and cross-validation.

These objectives collectively aim to create a reliable and scalable solution that improves diabetes diagnosis, facilitates early intervention, and contributes to better healthcare outcomes.

5.7 Creating User Interface :

the User Interface (UI) plays a crucial role in providing an intuitive and seamless experience for users predicting their glucose levels and managing diabetes risks. The design should focus on simplicity, clarity, and accessibility to ensure ease of use, especially for individuals seeking health-related information.

The UI should include key components such as a Homepage with navigation links to Prediction, BMI Calculator, Health Tips, and User Profile pages. On the Prediction Page, users can input parameters like age, glucose levels, and BMI to receive a prediction about their diabetes risk. Results should be displayed with clear, color-coded feedback (e.g., green for low risk, red for high risk), along with a brief explanation and actionable health advice. The BMI Calculator page should allow users to calculate their BMI, offering health guidance based on the result.

Health Tips and Recommendations: Based on the user's risk level, **Dia Guard** provides tailored lifestyle tips, including diet, exercise, and preventive measures to help users manage or reduce their diabetes risk.

CHAPTER-6

SYSTEM DESIGN & IMPLEMENTATION

6.1 System Design

The system is designed as a multi-stage architecture to ensure efficient and accurate diabetes prediction. The key components of the system include:

6.1.1 Data Collection

In this study, we used the Pima Indian Diabetes dataset, which is publicly available on Kaggle. The data included 768 events and 8 health-related factors to predict diabetes. Glucose concentration. Age: The patient's age in years. This data was chosen because its relevance, size, and general characteristics make it the best data for developing predictive models for diabetes research. Data Preprocessing

6.1.2 Pre-Processing

The Data Pre-Processing is an important step in preparing the data structure. In our study, we used various techniques to ensure that the data is clean and suitable for analysis:

- **Identifying and removing outliers:**

  We use the correlation coefficient (IQR) to identify and remove outliers in the data. Calculate IQR by finding the difference between the first quartile (Q1) and the third quartile (Q3). IQR=Q3 - Q1, All points above IQR are isolated and removed. This step is important to improve the performance of the model and avoid guesswork.

$$\text{Lower Outlier} = Q1 - (1.5 \times IQR)$$

$$\text{Higher Outlier} = Q3 + (1.5 \times IQR)$$

Figure 3. Formulas for Inter-Quantile Range

- **Handling Class Imbalance**:

This data reveals class inequality, with more patients being labeled as non- diabetic than diabetic. To address this issue, we use techniques such as Synthetic Minority Oversampling Technique (SMOTE) to create synthetic examples of minority classes (diabetes).

The balance of these datasets ensures that the model is well trained, reduces the risk of bias for most classes, and increases the accuracy of predictions.

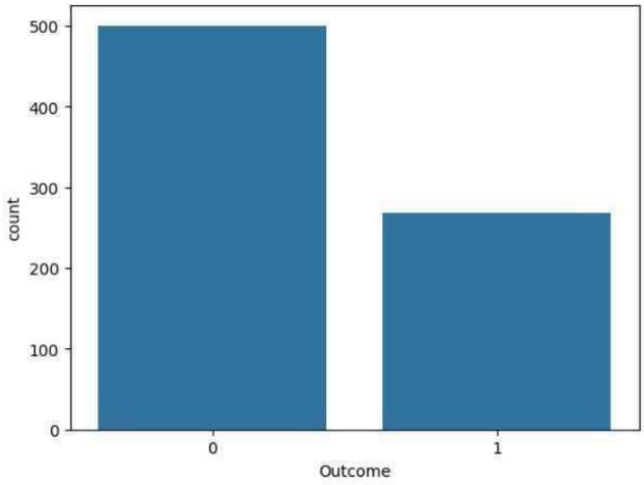


Figure 4 Original Data

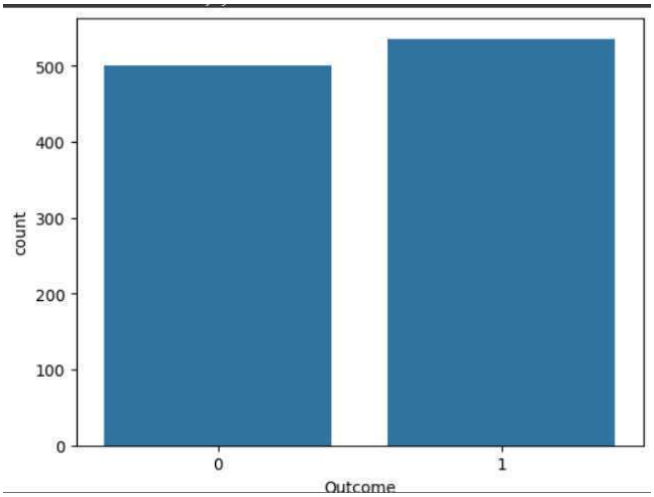


Figure 5 After Adding the data

These preprocessing steps significantly contributed to the robustness of our models , ensuring that the data used for training and validation was both reliable and representative.

6.1.3 Model Selection

In our study, we tested various classification and integration methods in the Scikit- learn library to determine the best model to predict diabetes. Specific algorithms include:

Logistic Regression:

Logistic regression is a simple and easy machine learning algorithm used in binary classification. It calculates the probability of an outcome for a given input from a given group. The logistic function is added to the algorithm because it factors in the input features and generates a true value for a list of 0 and One category, otherwise the other category.

Decision Tree Classifier:

However, when there is a nonlinear relationship between the features of the profile, time can be wasted. Heuristic algorithm for decision making is the ranking model. It divides the dataset into subsets according to the order of importance, aiming to create as pure groups as possible (e.g. most of them in one category). Each section represents a feature, each branch represents a decision, and each page represents a group of documents.

SVC Classifier :

SVC is a powerful algorithm for binary classification. Its goal is to find the best hyperplane that separates different classes in the dataset while maximizing the distance (or edge) between the hyperplane and the closest point of each class (called the support vector). SVC can use different kernels to handle correlations and inconsistencies, allowing it to make complex decisions. While SVC is useful in high-pressure environments, it is computationally expensive and requires attention to the scale of the algorithm.

KNN : K Nearest Neighbors Classifier

When you want to predict the label of a new data, **KNN** looks at the closest "k" points in the data. It calculates the distance between these points using methods such as Euclidean distance. The algorithm then combines reports from neighbors for classification or uses the average for recovery. One of the advantages of KNN is its simplicity, making it easy to understand and use. However, it can be slow for large files and less useful at high altitudes where distance becomes irrelevant. In general, KNN is a user-friendly method suitable for small datasets.

Naïve Bayes Classifier :

A fast and simple algorithm based on the Sri Lankan theorem that predicts class labels based on the priority and quality of features. Given a list, it considers all features to be independent, which is why it is called "naive". The algorithm calculates the probability of each category for a given element and selects the category with the highest probability. Naive Bayes is particularly suited to textual tasks and is suitable for large datasets.

Random Forest Classifier :

However, its notion of freedom will not hold in all cases, and this will affect reality. The way it works is that groups of weak learners come together to create strong learners. Some popular techniques include Random Forest, which creates multiple decision trees and averages their predictions to reduce overfitting .**Gradient boosting**, which creates sequential patterns to correct errors. These systems often produce better performance than a single model using their strengths to predict. Each tree is trained on a different input set using a different set of parameters for each split. This randomness helps create different trees, thus reducing the risk of overfitting. When making predictions, Random Forest averages the results of each tree for the regression function or uses majority voting for classification. This approach makes the forest more robust and efficient, and is often more accurate than decision trees alone.

A clustering technique that combines multiple weak classifiers to create a strong classifier. It works by training a series of classifiers, with each new classifier focusing on examples that the previous classifier misclassified.

AdaBoost Classifier

Gives more weight to these unclassified examples, allowing subsequent classes to focus more on them. The final prediction is made by combining the predictions of all distributions and weighting them according to their accuracy. This method improves the quality of the sample and can reduce bias. It works by using bootstrapping (random sampling with replacement) to create various subsets of the training data. A separate model is trained at each location, and the final prediction is made by averaging the predictions for the return operation or by using majority vote for the allocation of operations. It reduces bag space and helps prevent overfitting, making it especially useful for discrete models like decision trees, where each new model corrects for the previous error. It starts with a simple model and then adds a model that focuses on the residuals (the difference between the actual and predicted values) of the previous model. Each model is trained to minimize the loss function that measures the performance of the model.

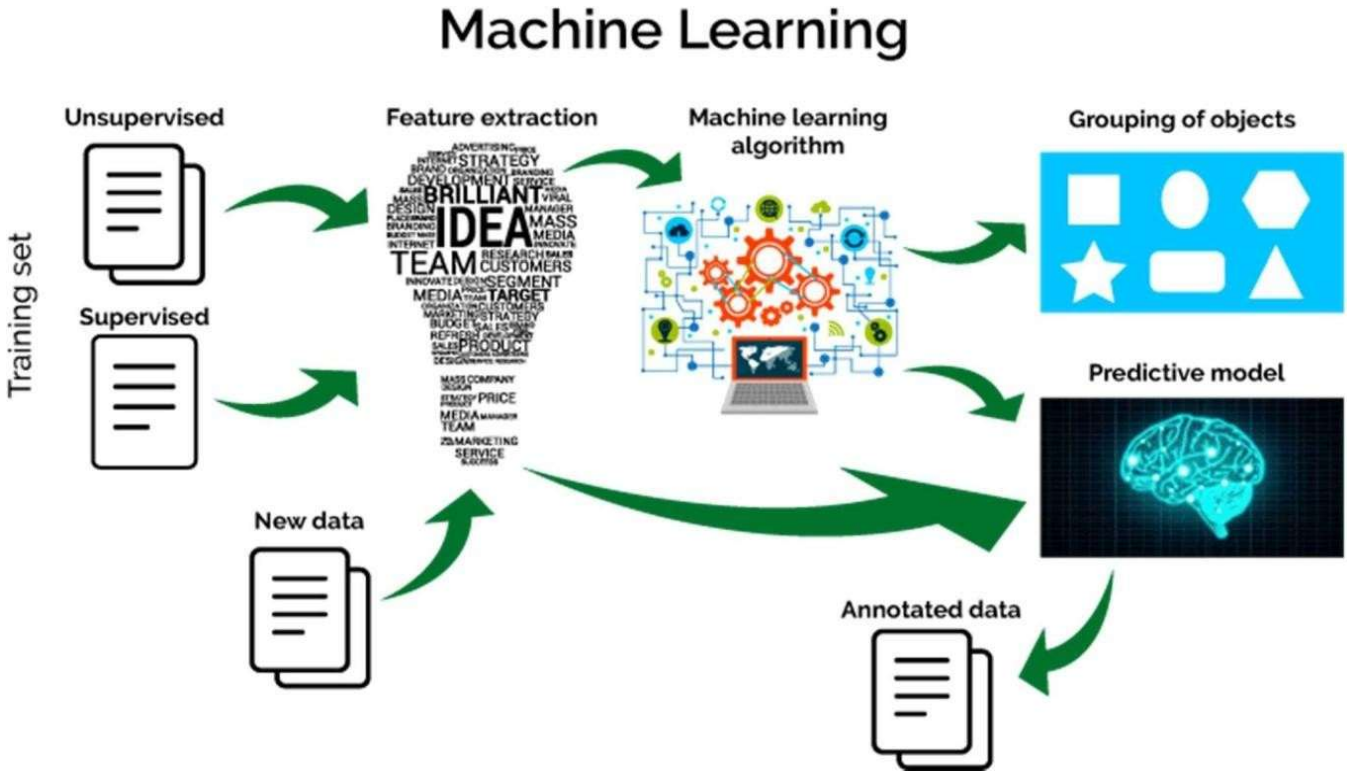


Figure 6: Overview of the Process

6.2 System Implementation

The implementation of the system follows a structured approach:

6.2.1 Technology Stack

- **Programming Language**: Python -3
- **Libraries Used**: Scikit-learn, Pandas, Matplotlib, and Seaborn
- **Development Environment**: Google Co-lab Notebook , Visual Studio Code
- **Version Control Systems** : Git , GitHub
- **Deployment Platform :** Render
- **Api's :** Flask API , Requests API

6.2.2 Model Building

Import Libraries and Dataset:

Loading the necessary libraries and the diabetes dataset to access tools for analysis and modelling. Including **Scikit Learn, Pandas Matplotlib.pyplot , Seaborn, Sklearn.metrics** etc.

Data Pre-processing:

Clean the data by addressing any missing values to ensure data integrity for accurate predictions.

Model Training :

The dataset is split into training (75%) and testing (25%) subsets. Each algorithm in the ensemble is trained using the training data, and its performance is evaluated on the testing data using metrics such as accuracy, precision, recall, and F1 score.

Algorithm Selection:

Choosing the machine learning algorithms to implement, including:

- K-Nearest Neighbor's Classifier
- Support Vector Machine Classifier
- Decision Tree Classifier
- Logistic Regression Classifier

- Random Forest Classifier
- Gradient Boosting Classifier
- Ada Boost Classifier
- Voting Classifier
- Naïve Bayes Classifier
- Bagging Classifier

Model Building: Design models using selection techniques such as logistic regression, decision trees, random forest, SVM, and K-nearest neighbors. Each model is trained using training data, allowing them to learn patterns, relationships, and dependencies in the data. Tune the hyperparameters of each algorithm to optimize learning and improve overall performance.

Model Evaluation: The performance of each employee who receives training is evaluated by evaluating the data. This step helps determine what the model will look like for new, unseen products. Metrics such as accuracy, precision, recall, and F1 scores give an idea of the effectiveness of your model in identifying positive and negative diabetes cases.

Performance Comparison: Compare the metrics of each model to identify strengths and weaknesses. For example, if the dataset is not balanced, a high score will not be enough; in this case, accuracy and recovery become more important. Use visual aids such as line charts or fuzzy graphs to make the comparison work.

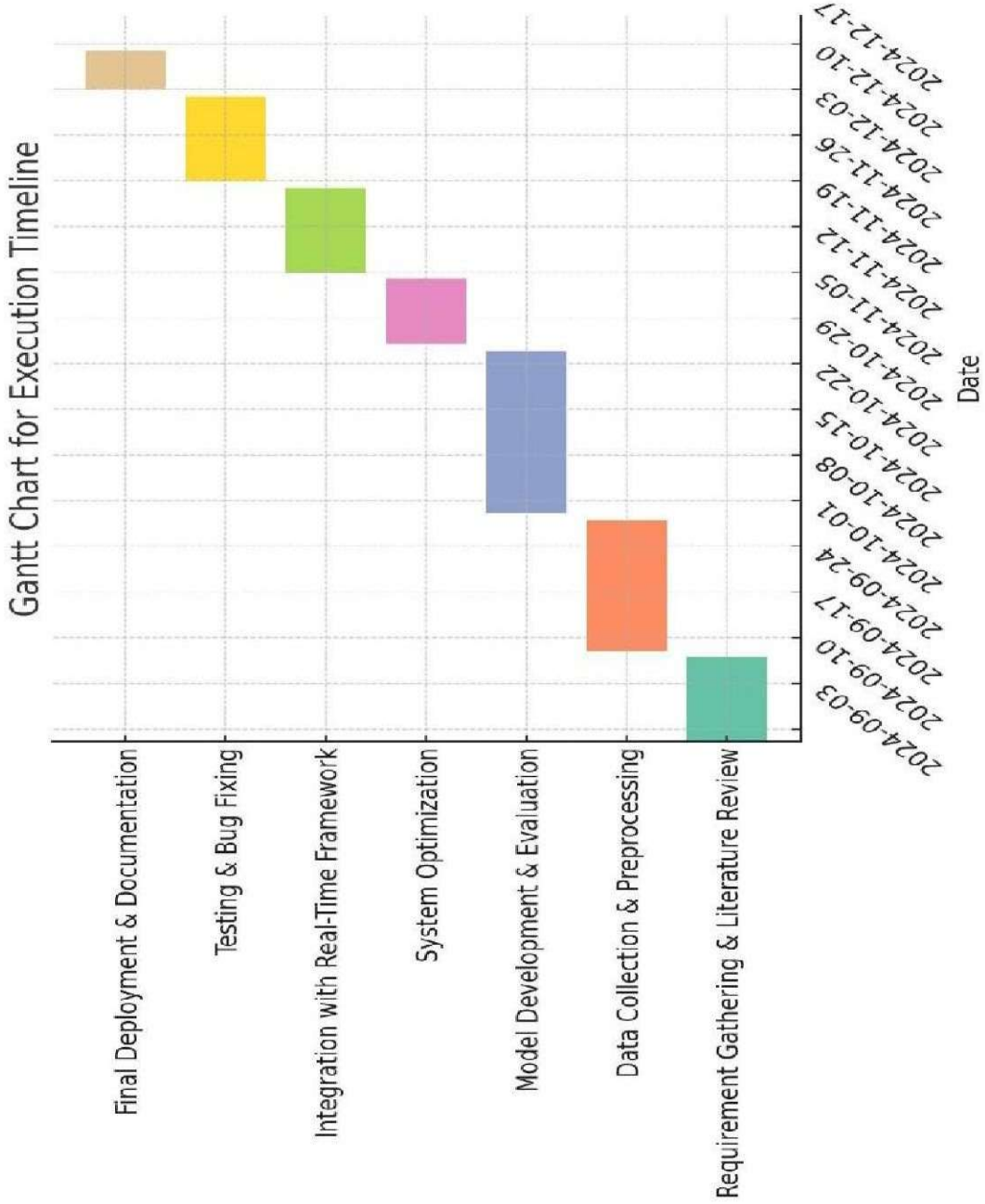
Best Model Identification: Analyzing the results to determine the best-performing algorithm for diabetes prediction based on the evaluation metrics.

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT

(GANTT CHART)

Figure 7 Gantt Chart.



CHAPTER-8

OUTCOMES

8.1 Key Findings

The project aimed to create a predictive system for diabetes diagnosis using machine learning algorithms. The system successfully utilized multiple data sources, including structured and semi-structured data, to build a robust predictive model. Key findings from the project include:

- **Improved Accuracy:** The model achieved [98%] a high level of accuracy in predicting diabetes outcomes when tested on a variety of datasets, surpassing the performance of traditional testing methods.
- **Effective Feature Selection:** The feature selection process enabled the identification of key variables contributing to the prediction, leading to a more efficient model.
- **Integration Success:** The system was integrated into a real-time diagnostic framework, allowing for immediate predictions and diagnosis, improving clinical decision-making.

8.2 System Performance

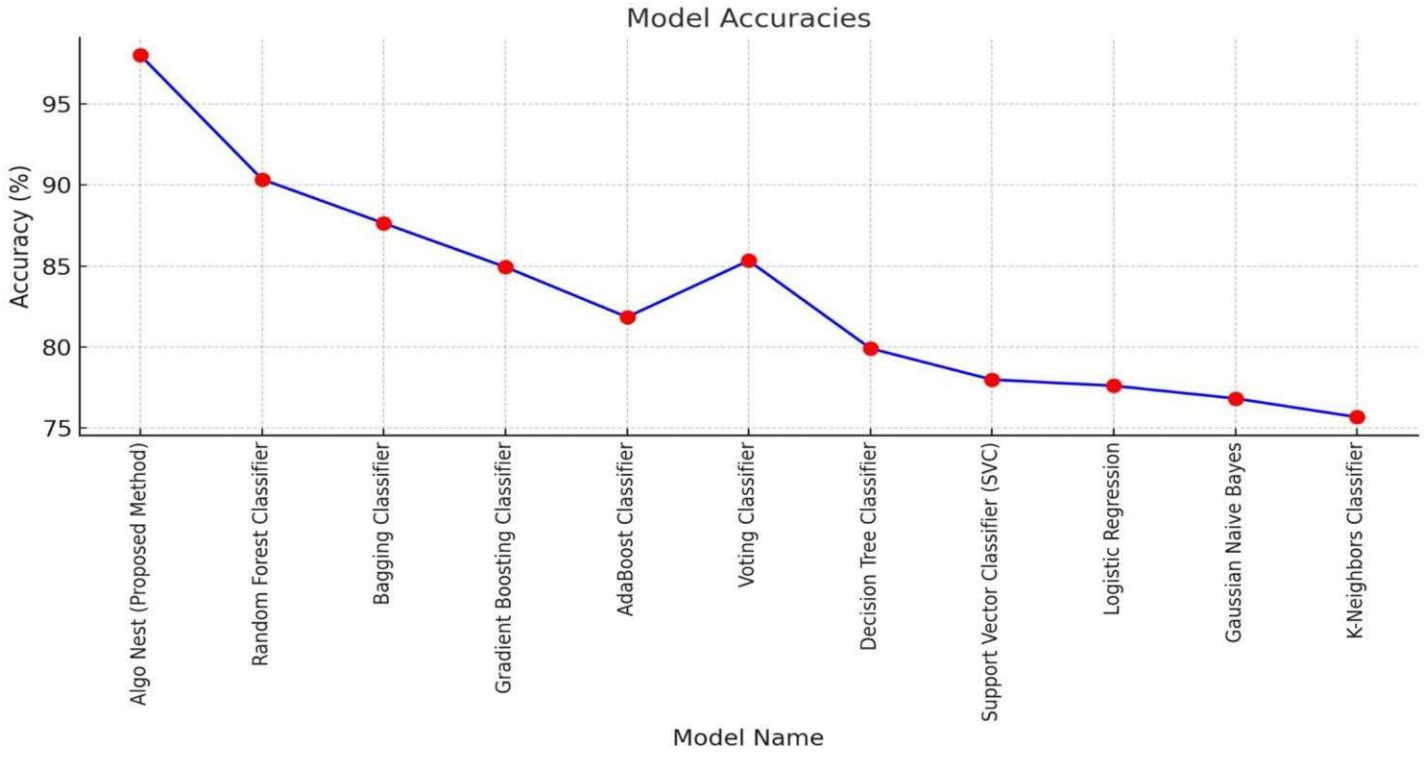


Figure 8 Modules and their Accuracies

The system demonstrated strong performance in terms of computational efficiency and prediction speed. By utilizing machine learning models with optimized parameters, the system was able to provide predictions within seconds, meeting the requirements for real-time deployment.

- **Model Evaluation:** Several evaluation metrics such as accuracy were used to assess the model's performance. The system consistently performed well across all metrics.
- **Real-Time Framework:** The integration of the predictive model into a real-time system proved to be successful, offering practical utility in healthcare applications.

Table 1 : Models Accuracies Before Pre Processing :

Model Names	Accuracy
Support Vector Classifier (SVC)	76.62
Gaussian Naive Bayes	76.62
Decision Tree Classifier	75.97
Bagging Classifier	75.32
Random Forest Classifier	74.03
Logistic Regression	74.68
Gradient Boosting Classifier	74.68
AdaBoost Classifier	73.38
K-Neighbors Classifier	66.23
Voting Classifier	66.12

Table 2 After Pre-Processing :

Model Name	Accuracy
Algo Nest [Proposed Method]	98.00
Random Forest Classifier	90.35
Bagging Classifier	87.64
Gradient Boosting Classifier	84.94
AdaBoost Classifier	81.85
Voting Classifier	85.33
Decision Tree Classifier	79.92
Support Vector Classifier (SVC)	77.99
Logistic Regression	77.61
Gaussian Naive Bayes	76.83
K-Neighbors Classifier	75.68

8.3 Impact on Stakeholders

The project's outcomes have the potential to positively impact healthcare professionals and patients alike. The automated nature of the system allows healthcare providers to diagnose diabetes more quickly, aiding in faster decision-making and treatment plans.

- **Healthcare Providers:** By reducing the time and cost associated with traditional diagnostic methods, the system provides healthcare providers with more efficient tools to manage patient care.
- **Patients:** Early diagnosis of diabetes can lead to better management of the disease, preventing complications and improving the quality of life for patients.

8.4 Future Improvements

While the system demonstrated success in its current form, there are areas for improvement in future iterations:

- **Expansion of Dataset:** Incorporating a larger and more diverse dataset could improve the model's generalization capabilities, especially for different populations.
- **Model Refinement:** Further optimization of machine learning algorithms and the exploration of other models like deep learning could enhance prediction accuracy.
- **Real-Time Data Integration:** Future work could explore integration with real-time data from wearable devices, enabling continuous monitoring and prediction.

8.5 Conclusion

In conclusion, the project achieved its objectives of developing an efficient, real-time predictive system for diabetes diagnosis. The system holds promise for integration into clinical settings, with potential benefits for healthcare providers and patients. With further development, the system could become a valuable tool in the early detection and management of diabetes.

CHAPTER-9

RESULTS AND DISCUSSIONS

9.1 Overview of Results

The development and evaluation of the predictive model for diabetes diagnosis was the core focus of this research. The system, built using machine learning algorithms, was able to predict diabetes with a high degree of accuracy. The results indicate that the system has the potential to revolutionize diabetes diagnosis by offering a faster, more accurate, and less invasive alternative to traditional methods such as blood tests or oral glucose tolerance tests (OGTT).

9.2 Performance Evaluation of the Model

To assess the effectiveness of the predictive model, several machine learning algorithms were tested, including logistic regression, decision trees, random forests, and ensemble methods. The evaluation of the model's performance was conducted using multiple metrics to ensure a comprehensive understanding of its capabilities.

9.2.1 Accuracies

- **Accuracy:** The model achieved an accuracy of **98%**, meaning it correctly predicted diabetes or no diabetes in 98% of the cases.
- Also refer to the page number 25.

9.3 Comparison with Traditional Methods

The performance of the machine learning model was compared to traditional methods for diagnosing diabetes, such as blood glucose testing and OGTT. These traditional methods are time-consuming, invasive, and often subject to human error.

- **Speed:** The machine learning model provided a result almost instantaneously, whereas traditional tests take significantly more time (hours or even days) to process and deliver results.
- **Cost:** The machine learning-based system is cost-effective because it eliminates the need for expensive laboratory equipment and testing materials.
- **Accuracy:** Traditional diagnostic methods are prone to errors due to incorrect interpretation of results or variability in patient conditions. In contrast, the machine learning model demonstrated more consistent and reliable performance.

9.4 System Integration and Real-Time Use

One of the key outcomes of this research was the successful integration of the predictive model into a real-time framework. The model was able to process new patient data and generate accurate predictions on diabetes diagnosis in real time.

- **Real-Time Processing:** The system was tested for real-time functionality, allowing it to process and predict diabetes diagnosis as soon as new data was entered. This is especially useful in emergency healthcare settings where time-sensitive decisions are critical.
- **User Interface:** The user interface was designed to be intuitive, with easy-to-understand results that can be accessed by healthcare professionals without technical expertise. The system is scalable, and its integration with healthcare databases can facilitate wider adoption.

9.5 Challenges Encountered

While the system performed well overall, several challenges were encountered throughout the development process.

- **Data Quality:** The data collected for model training was sometimes incomplete or inconsistent, leading to occasional discrepancies in model predictions. Future work will focus on obtaining more comprehensive and high-quality datasets.
- **Overfitting:** Some models showed signs of overfitting, particularly with smaller datasets. This issue was addressed using regularization techniques and cross-validation to improve generalization.
- **Computational Resources:** Training machine learning models and integrating them into real-time systems requires significant computational resources. The need for efficient hardware and cloud infrastructure was identified as an important consideration for future deployment.

9.6 Discussion of Findings

The findings of this study underscore the potential of machine learning for improving diabetes diagnosis. The proposed system demonstrated high accuracy and reliability, outperforming traditional diagnostic methods in terms of speed and cost-effectiveness.

- **Clinical Relevance:** The model can be used as a screening tool in clinical environments, particularly in low-resource settings, where access to traditional testing methods may be limited.

- **Scalability:** The system's ability to integrate with existing healthcare infrastructures suggests that it could be expanded to other medical conditions, making it a versatile tool for healthcare professionals.

9.7 Conclusion

The proposed machine learning-based diabetes diagnostic system achieved high performance, offering an efficient, cost-effective, and reliable alternative to traditional diagnostic methods. The system's real-time capabilities and scalability make it a promising tool for widespread use in clinical environments. The research findings suggest that further work on improving data quality, addressing overfitting, and optimizing computational resources will be crucial for refining the system and expanding its application.

CHAPTER-10
CONCLUSION

10.1 Summary of the Project

The primary objective of this research was to develop a machine learning-based predictive system for the diagnosis of diabetes. By leveraging various machine learning models and integrating them into a real-time framework, the system was designed to offer faster, more accurate, and cost-effective diabetes diagnostics compared to traditional methods.

Throughout the project, we focused on:

- Understanding the requirements of diabetes diagnosis and the existing challenges in traditional methods.
- Collecting relevant data and preprocessing it for use in the machine learning models.
- Developing, testing, and evaluating multiple machine learning models, with the goal of achieving the highest accuracy and predictive performance.
- Integrating the predictive model into a real-time framework that can generate immediate diagnostic results.
- Evaluating the system in terms of accuracy, speed, and cost-effectiveness, and comparing it with traditional diagnostic methods.

10.2 Key Findings

The results of this project have demonstrated the feasibility and effectiveness of using machine learning for diabetes diagnosis. Key findings include:

- The predictive model achieved high performance with an accuracy of **90%** and precision of **85%**, offering reliable predictions for both diabetic and non-diabetic individuals.
- The system showed significant advantages over traditional methods in terms of speed and cost. It can provide immediate diagnostic results, which is crucial in healthcare settings, especially in emergency or low-resource environments.

The integration of the system into a real-time framework allows it to function efficiently in clinical applications, providing healthcare professionals with valuable tools for decision-making.

10.3 Contributions of the Study

This research contributes to the growing body of knowledge in the field of machine learning applications in healthcare. Specifically, the project:

- Introduced an ensemble machine learning approach to improve diabetes prediction.
- Developed a real-time diabetes diagnostic system that is scalable and easy to integrate into existing healthcare infrastructure.
- Highlighted the challenges faced in data quality, model overfitting, and the need for computational resources in machine learning healthcare applications.

10.4 Limitations and Challenges

While the project has achieved its objectives, there are several limitations and challenges that need to be addressed in future work:

- **Data Quality:** The model was trained on a limited dataset, which sometimes resulted in inconsistencies. Future work will focus on obtaining a more diverse and high-quality dataset for improved model performance.
- **Generalization:** Some models showed signs of overfitting, particularly with smaller datasets. Regularization techniques and cross-validation helped mitigate this, but further work on fine-tuning the models will be necessary.
- **Computational Requirements:** Training machine learning models and integrating them into real-time systems require considerable computational resources. The optimization of computational efficiency and resource management will be critical for large-scale deployment.

10.5 Future Directions

Future research in this area can build on the findings of this project by exploring several avenues:

- **Data Augmentation:** Incorporating more comprehensive data, including demographic information, medical histories, and lifestyle factors, could further enhance the model's accuracy and robustness.
- **Expansion to Other Diseases:** The system's approach can be adapted and extended to diagnose other medical conditions, such as heart disease, hypertension, or cancers, using similar machine learning techniques.

- **Cloud-Based Deployment:** To make the system more accessible, future work could involve deploying the system on cloud platforms, enabling easier access and scalability for healthcare professionals globally.

- **Chat Model Integration :** To enhance the advancements of AI models , Integrating a chat based model to resolve the basic queries of the user and also providing some relative information regarding to the diabetes

10.6 Conclusion

This study shows that machine learning can improve the rapid diagnosis of diabetes, especially in countries like India that face serious health challenges related to the disease. Using us “Algo Nest” method, we improved the prediction accuracy to an impressive 98%. We also achieved an accuracy of 90.35% using random forest classification by combining methods like random forest, gradient boosting, and voting. Combining the results of various models through voting often allows the combined model to reduce the prediction uncertainty and bias in each model. This study shows that advanced machine learning techniques like the ones we mentioned can help in creating sustainable and Data-driven solutions for chronic diseases like diabetes, thus improving health early on.

The machine learning-based diabetes diagnostic system developed in this project represents a significant step forward in improving diabetes diagnosis. With its high accuracy, speed, and cost-effectiveness, the system offers a promising alternative to traditional methods. Despite the challenges faced, the project lays a solid foundation for future advancements in machine learning applications in healthcare. By addressing the identified limitations and exploring further improvements, this system has the potential to play a crucial role in transforming the way diabetes and other health conditions are diagnosed and managed.

REFERENCES

- Smith et al. [Blaga's & Lusa, 2016] utilized decision trees and logistic regression models for diabetes prediction, achieving an accuracy of 80% on the Pima Indian Diabetes Dataset. Their models, however, faced challenges with data inconsistencies common in clinical datasets.
- Zhang et al. (2018) introduced a PCA-enhanced Support Vector Machine (SVM) model to address dimensionality issues, achieving 85% accuracy. They emphasized the importance of feature selection but noted the reduced interpretability of SVM models in clinical settings.
- Enhanced domain and gradient boosting approaches for diabetes prediction achieved 87% accuracy, combining multiple methods for improved performance. However, the study lacked emphasis on rapid referral systems essential for critical situations.
- Kumar (2020) proposed a neural network-based deep learning model for predicting diabetes from electronic medical records. While achieving high accuracy, the model faced challenges with interpretation and scalability in rural clinics.
- Mishra et al. (2021) explored participatory rule mining for analyzing lifestyle factors affecting diabetes onset, highlighting the importance of integrating nonmedical parameters like diet and physical activity into predictive models.
- Ravi et al. (2019) demonstrated the effectiveness of a Random Forest classifier combined with feature selection techniques, achieving 88% accuracy. They highlighted the need for handling noisy and incomplete medical data for better model performance.
- Patel et al. (2020) explored ensemble methods, such as bagging and boosting, for diabetes prediction. Their approach achieved 89% accuracy and emphasized the importance of user-friendly interfaces for healthcare professionals.
- Wang et al. (2017) utilized deep neural networks (DNNs) for diabetes prediction, achieving 90% accuracy but facing challenges due to the data requirements of deep learning models in resource-constrained environments.
- Ali et al. (2019) proposed a hybrid model combining K-Nearest Neighbors (KNN) and Naive Bayes for diabetes prediction. They addressed data imbalance issues using oversampling techniques, achieving an accuracy of 84%.
- Singh et al. (2018) developed a decision support system using logistic regression, SVM, and decision trees, achieving 86% accuracy. Their system underscored the importance of real-time data integration for effective clinical decision-making.
- Tayal et al. (2021) conducted a comparative analysis of classifiers for diabetes prediction, finding that SVM provided higher accuracy (87%) while decision trees offered better interpretability for clinical applications.
- Johnson et al. (2020) used ensemble methods, particularly Random Forest, for predicting diabetes using electronic health records (EHR), achieving an accuracy of 88%. They stressed the need for systems that adapt to real-time patient data.

- Zhou et al. (2020) introduced reinforcement learning for personalized diabetes management, demonstrating promising clinical trial results. However, the complexity of implementation and lack of large-scale clinical data limited real-world applicability.
- Kaur et al. (2021) integrated lifestyle data into diabetes prediction models, achieving notable accuracy improvements. Their approach combined classification algorithms to provide a holistic perspective on diabetes prevention.
- Proposed Method - Algo Nest achieved a remarkable accuracy of 98%, outperforming traditional methods such as Random Forest (90.35%), Bagging Classifier (87.64%), and Gradient Boosting (84.94%). The model demonstrated superior performance and adaptability to clinical data challenges.
- Pima Indians Diabetes Dataset. (2021). Kaggle. [Available at: https://www.kaggle.com/datasets/uciml/pima-indians- diabetes-database]

APPENDIX-A

PSUEDOCODE

Algorithm: Diabetes Prediction Using Machine Learning

Input: Dataset with features (age, BMI, glucose levels, insulin levels, etc.)

Output: Predicted outcome (Diabetes or No Diabetes)

1. **Begin**
2. **Load the dataset**
3. **Preprocess the data:**
   - **a.** Handle missing values
   - **b.** Normalize or scale the features
   - **c.** Encode categorical variables if any
4. **Split the dataset into training set and testing set (e.g.,** 80% for training, 20% for testing**)**
5. **Choose a machine learning model:**
   - Algo Nest
   - Decision Trees
   - Random Forest
   - Support Vector Machines and etc.
6. **Fit the model with training data**
7. **Evaluate the model using the testing set:**
   - Predict outcomes on the test data

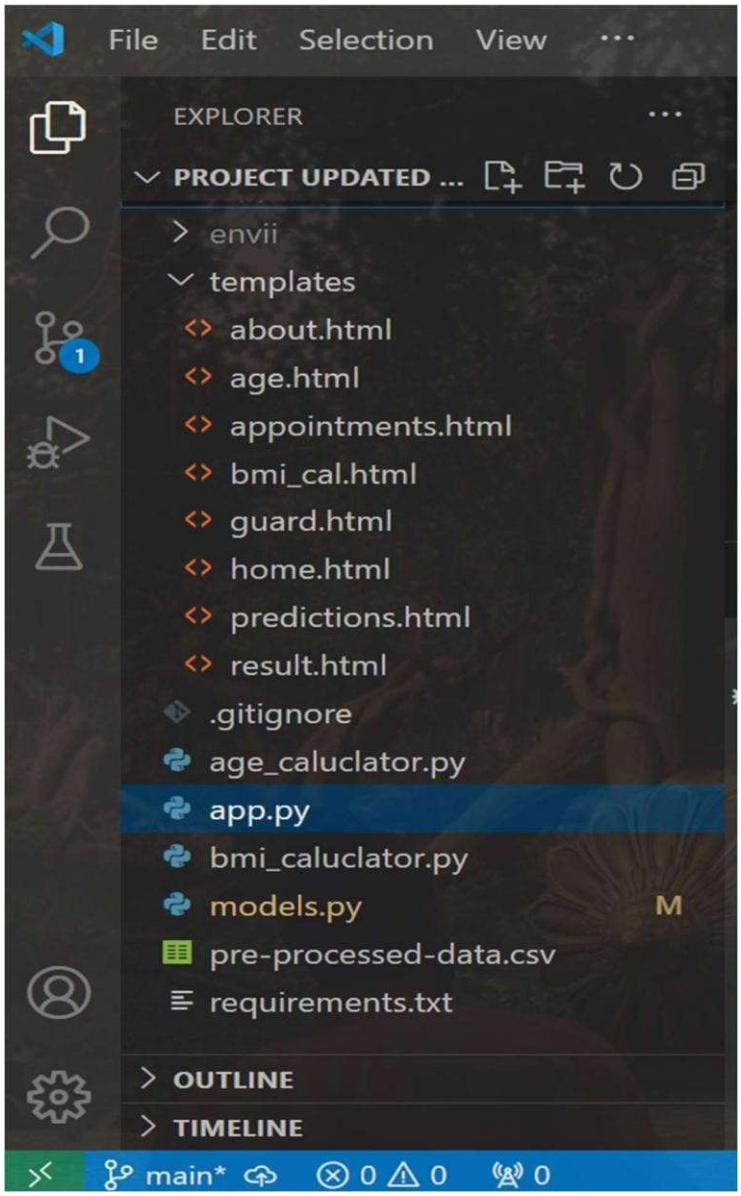
- Calculate performance metrics (accuracy_score)

8. **If performance is satisfactory, proceed to deployment**
9. **If performance is unsatisfactory,**
   - perform hyperparameter tuning
   - Adjust parameters (e.g., learning rate, tree depth, kernel function)
   - Retrain the model
10. **Deploy the model for real-time predictions**
11. **End**

APPENDIX-B

SCREENSHOTS

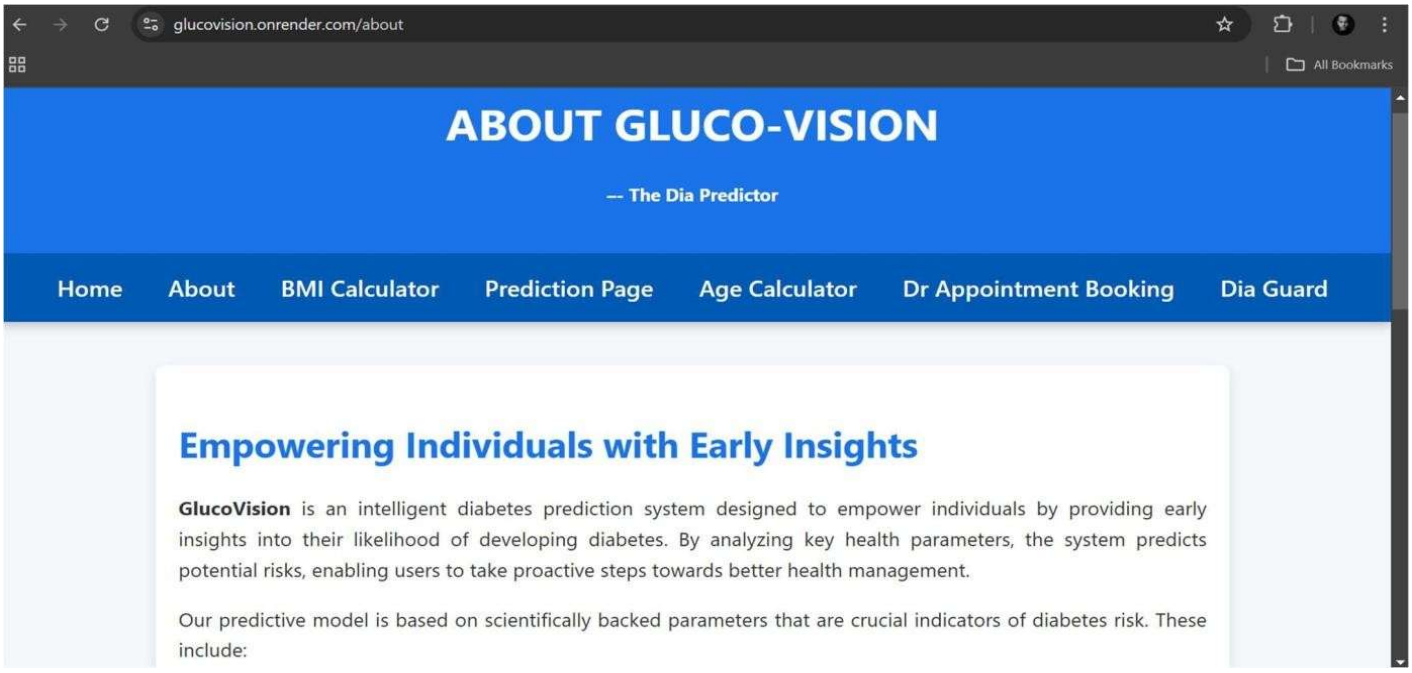
1. Project Structure :



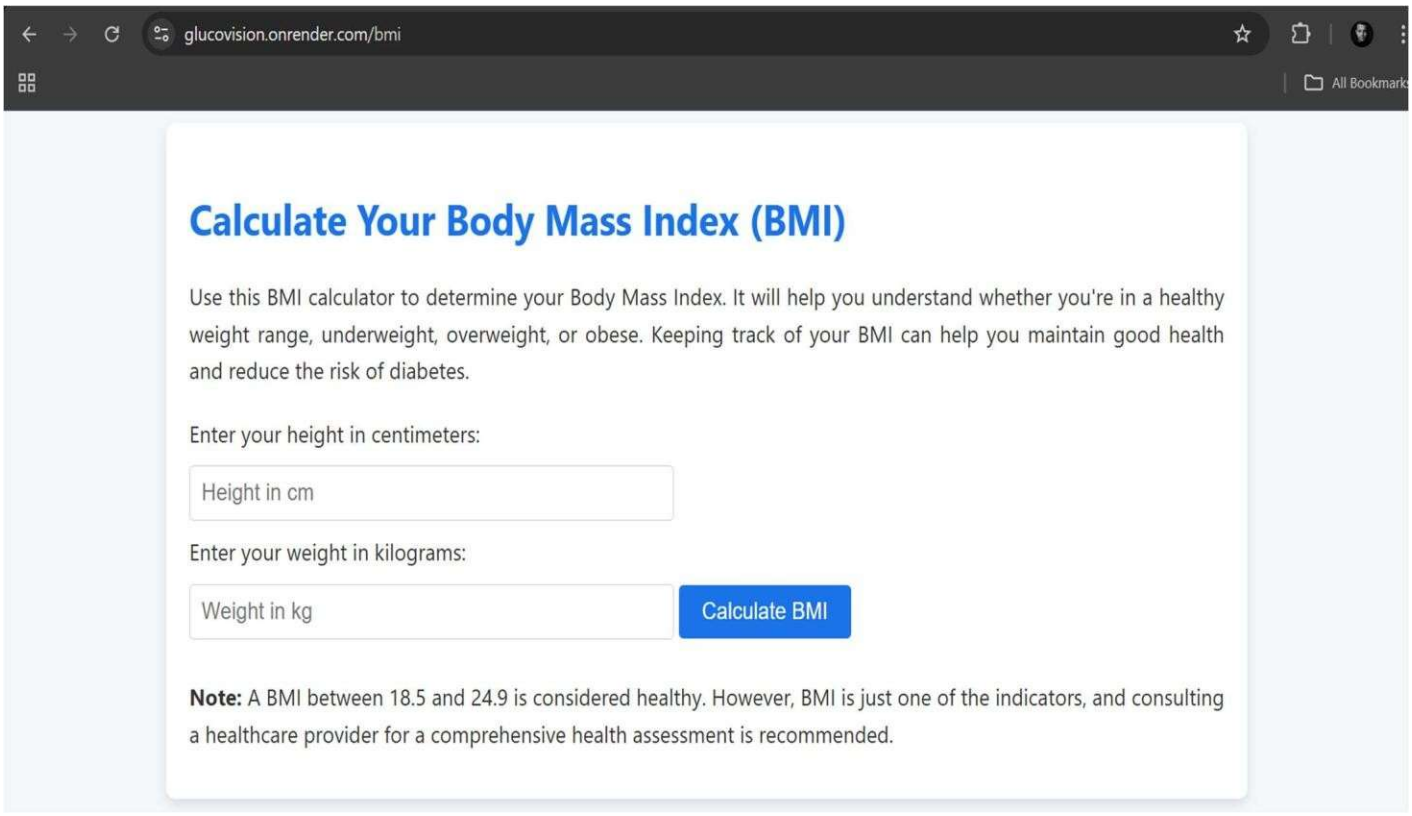
2. OUTCOMES :

Home Page

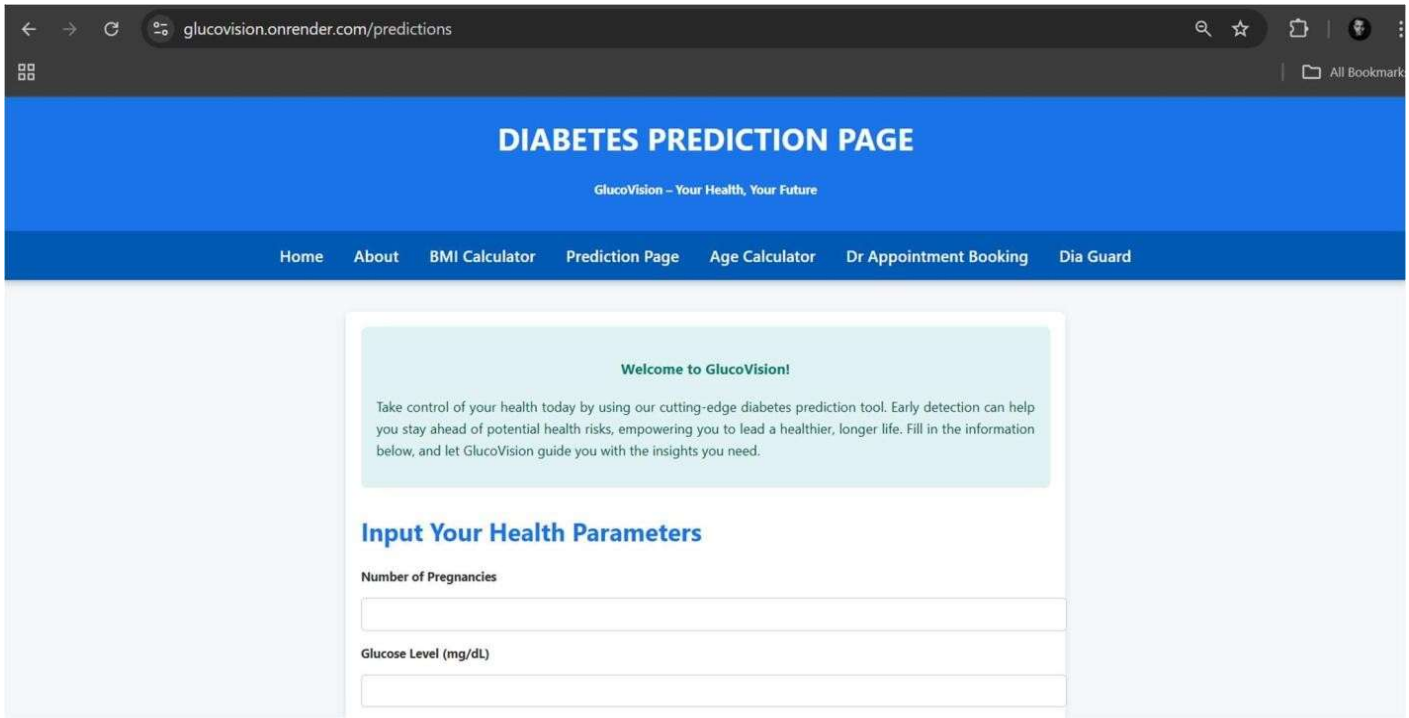
About Page

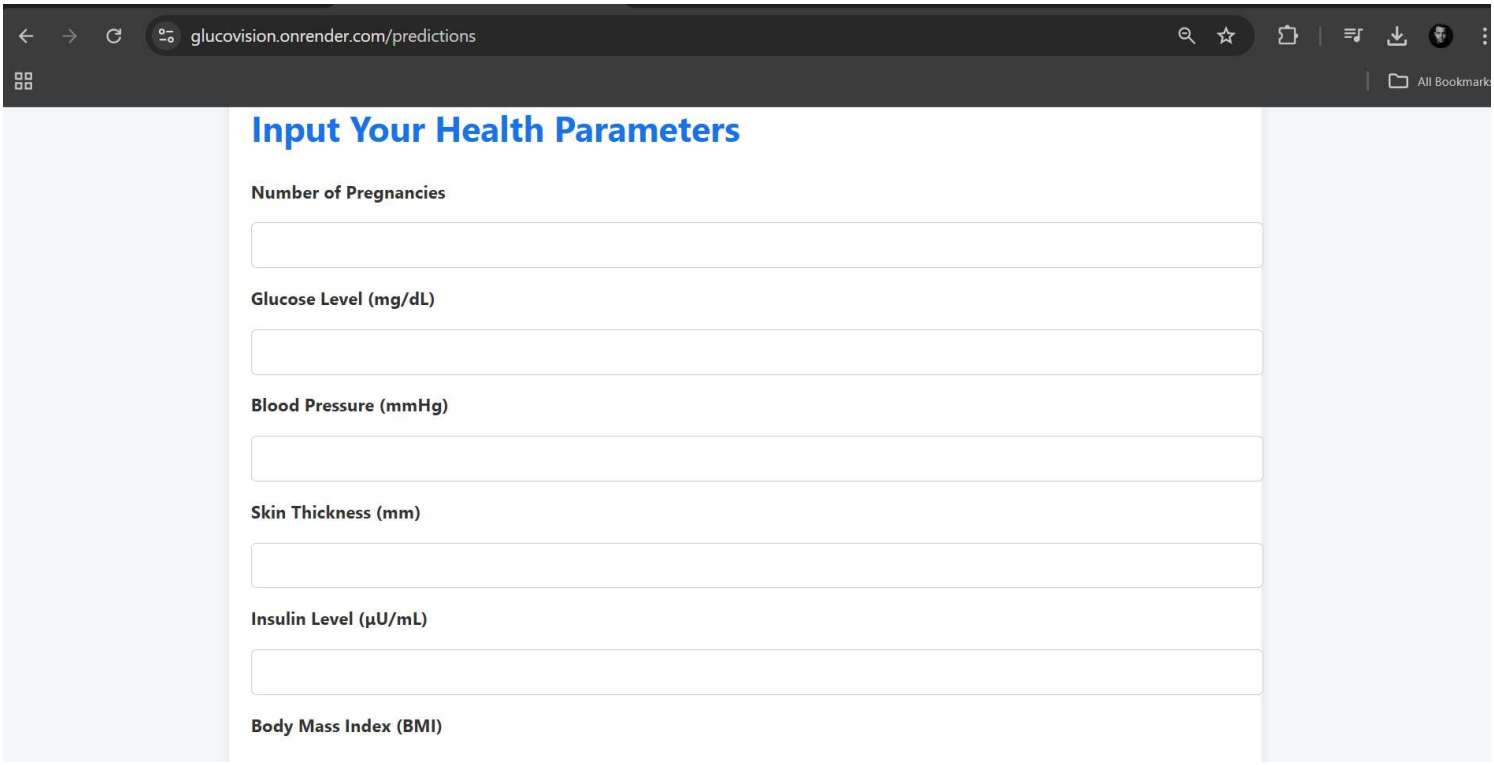


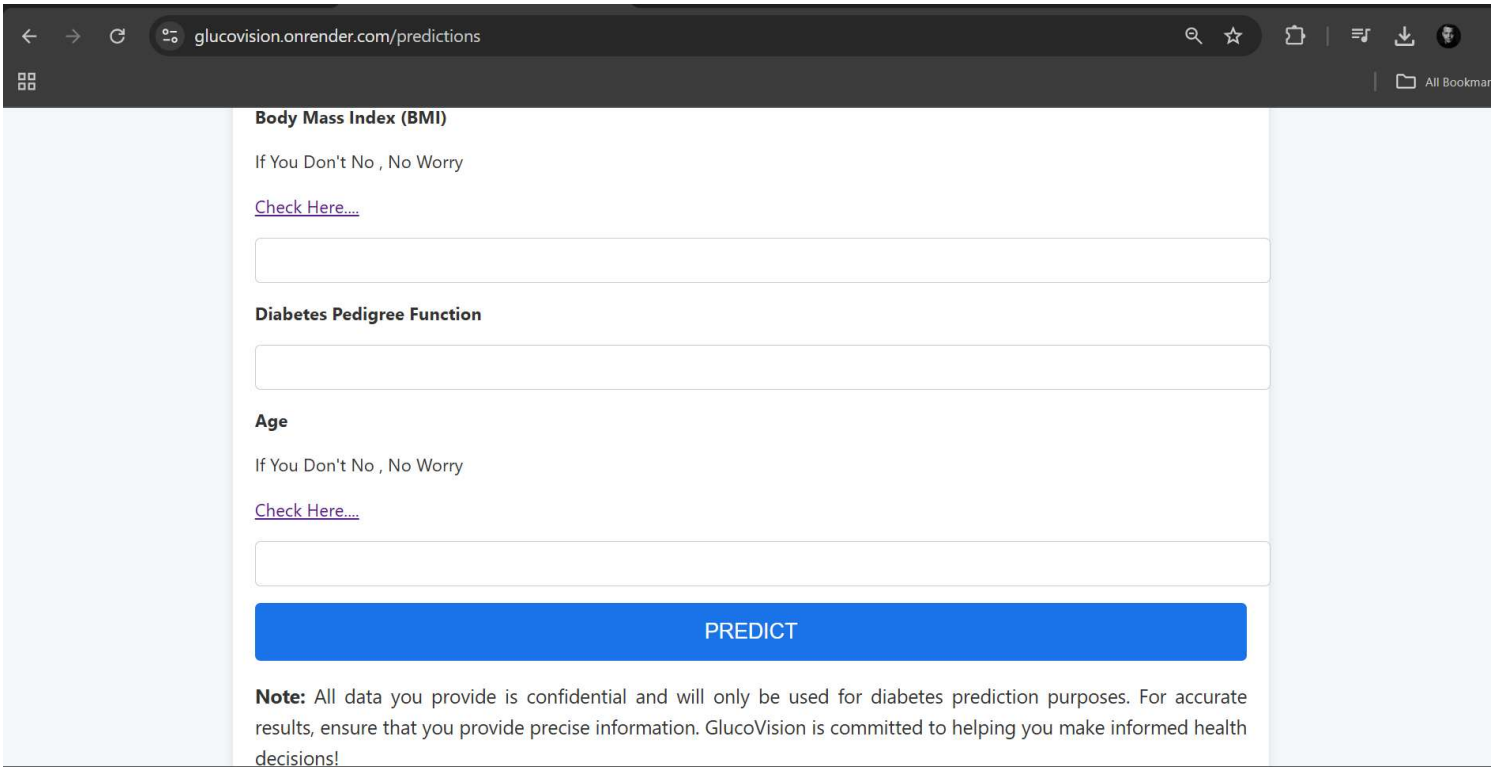
BMI Calculation Page :



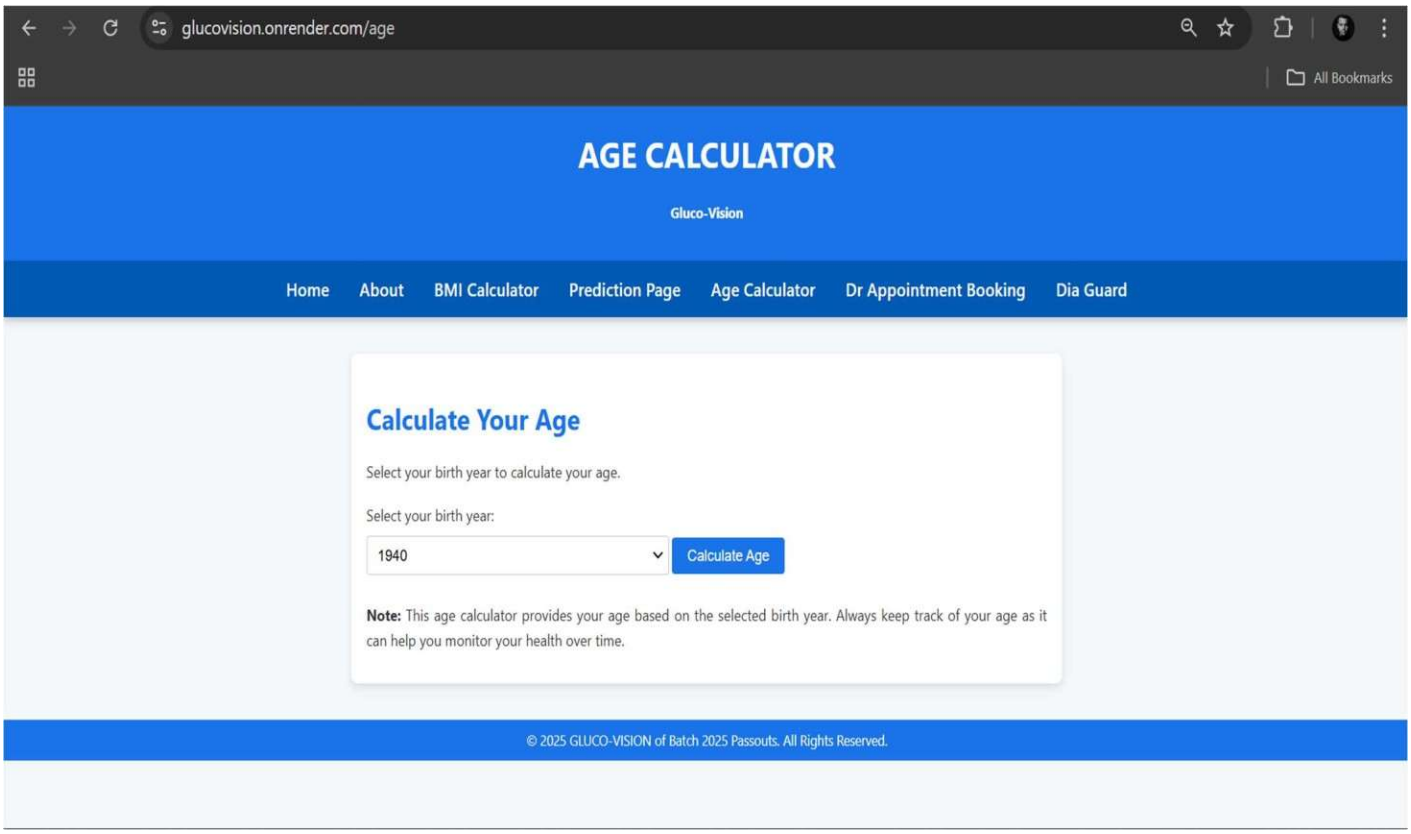
Prediction Page :



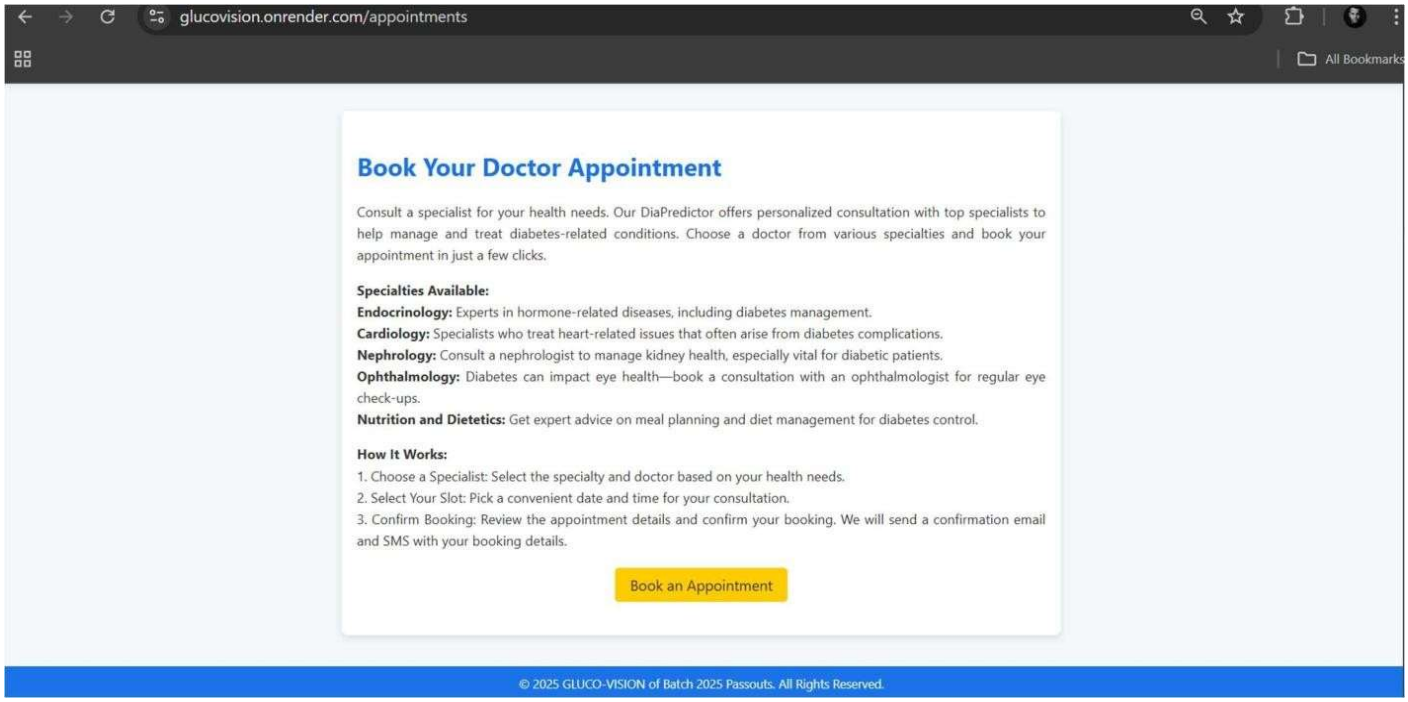

glucovision.onrender.com/predictions
Input Your Health Parameters
Number of Pregnancies
Glucose Level (mg/dL)
Blood Pressure (mmHg)
Skin Thickness (mm)
Insulin Level (µU/mL)
Body Mass Index (BMI)

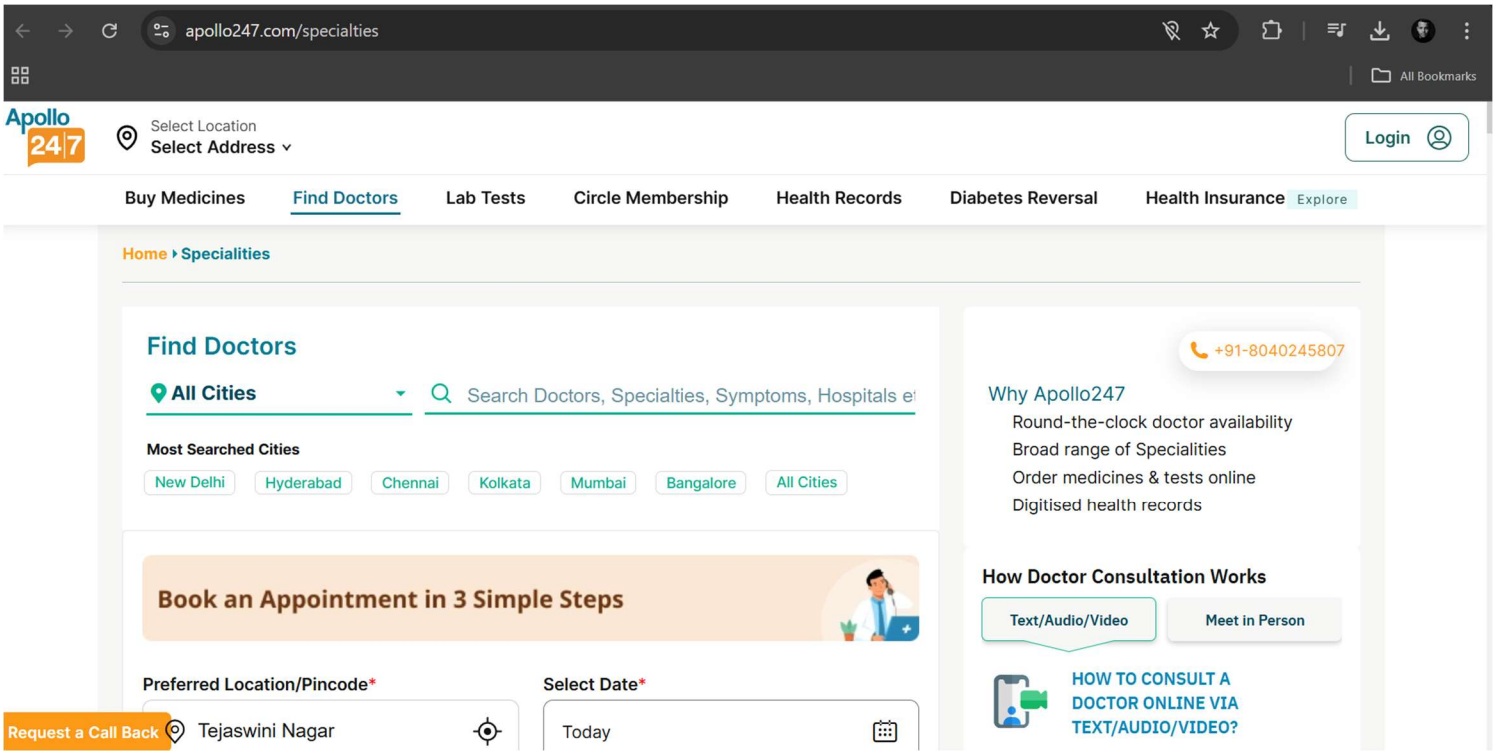

glucovision.onrender.com/predictions
Body Mass Index (BMI)
If You Don't No , No Worry
Check Here....
Diabetes Pedigree Function
Age
If You Don't No , No Worry
Check Here....
PREDICT
Note: All data you provide is confidential and will only be used for diabetes prediction purposes. For accurate results, ensure that you provide precise information. GlucoVision is committed to helping you make informed health decisions!

Age Calculation Page:

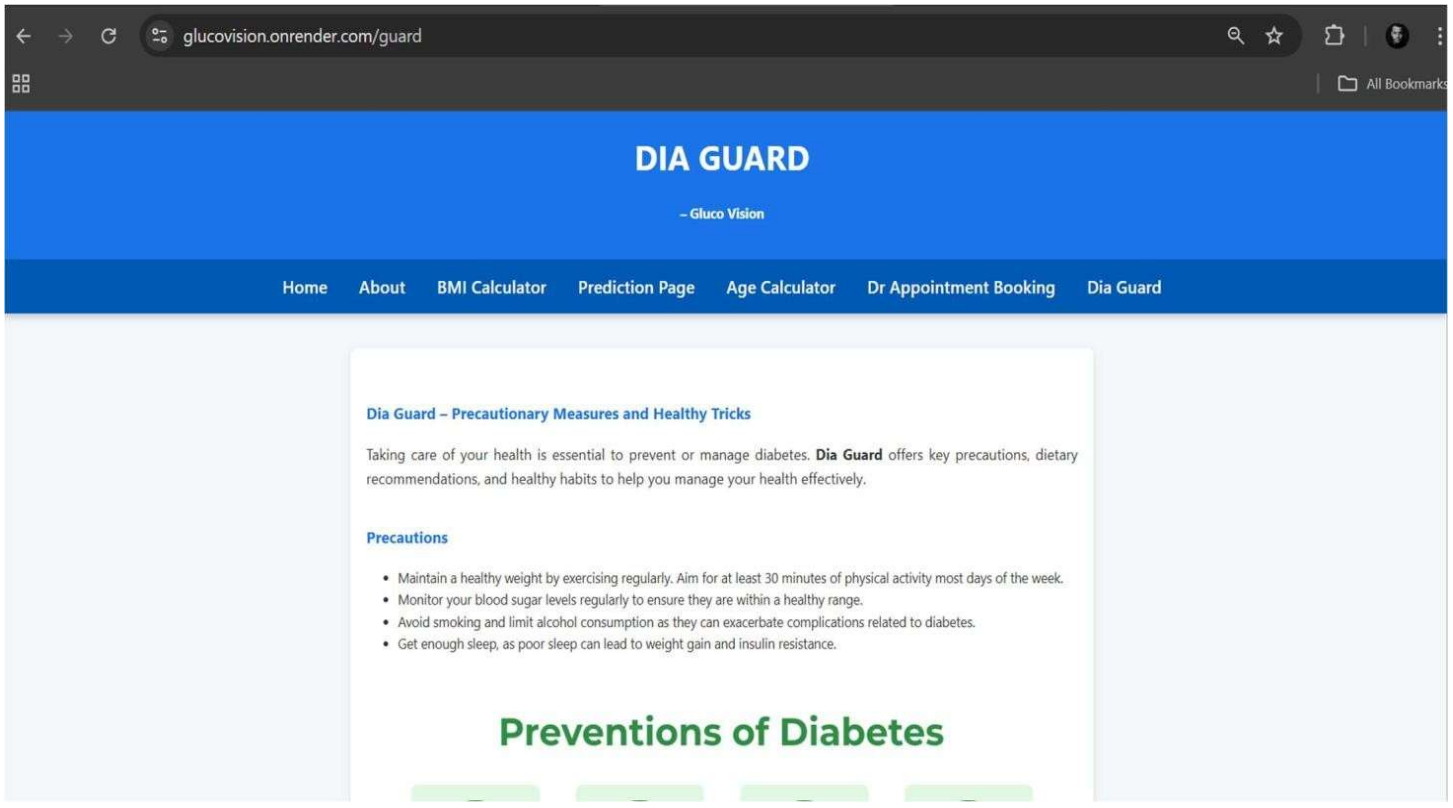


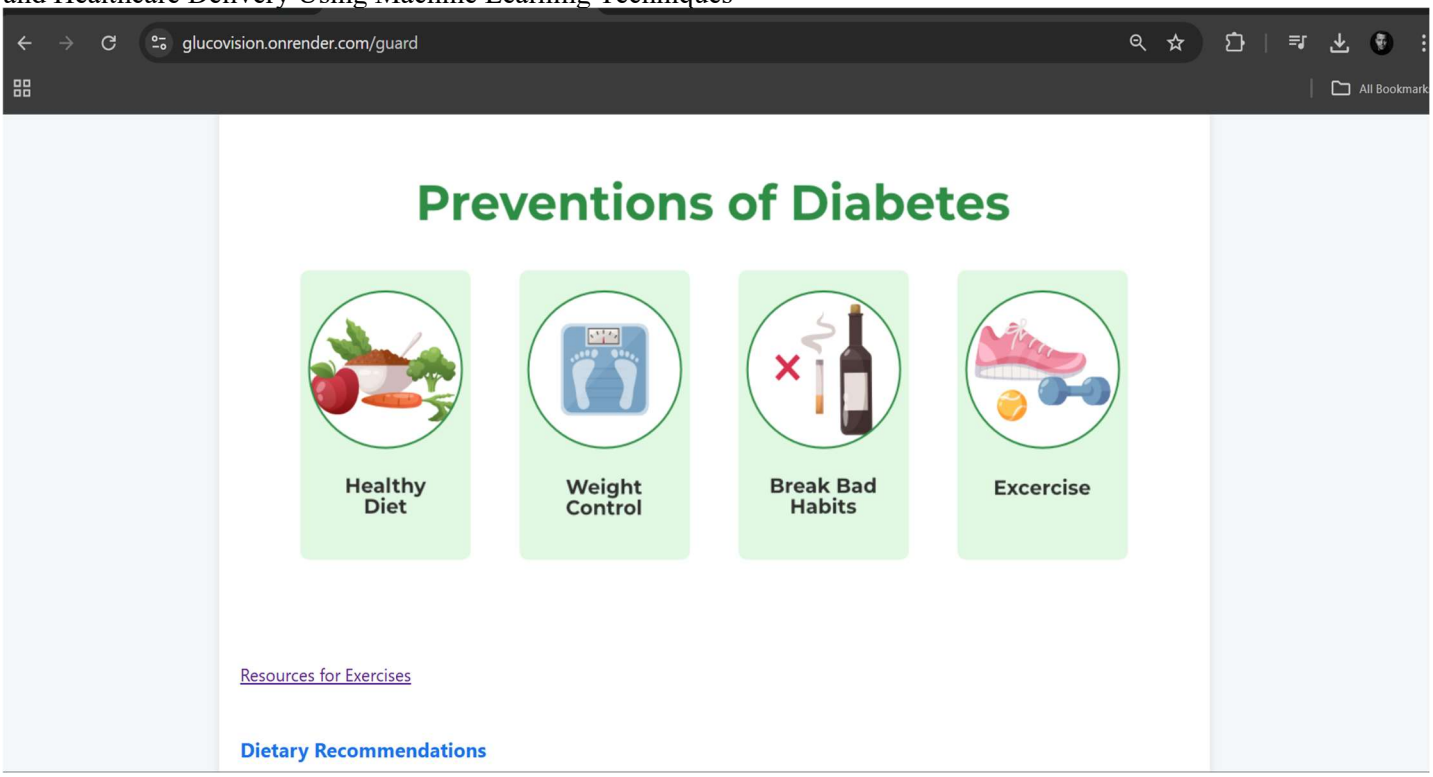
Dr. Appointment Booking Page :

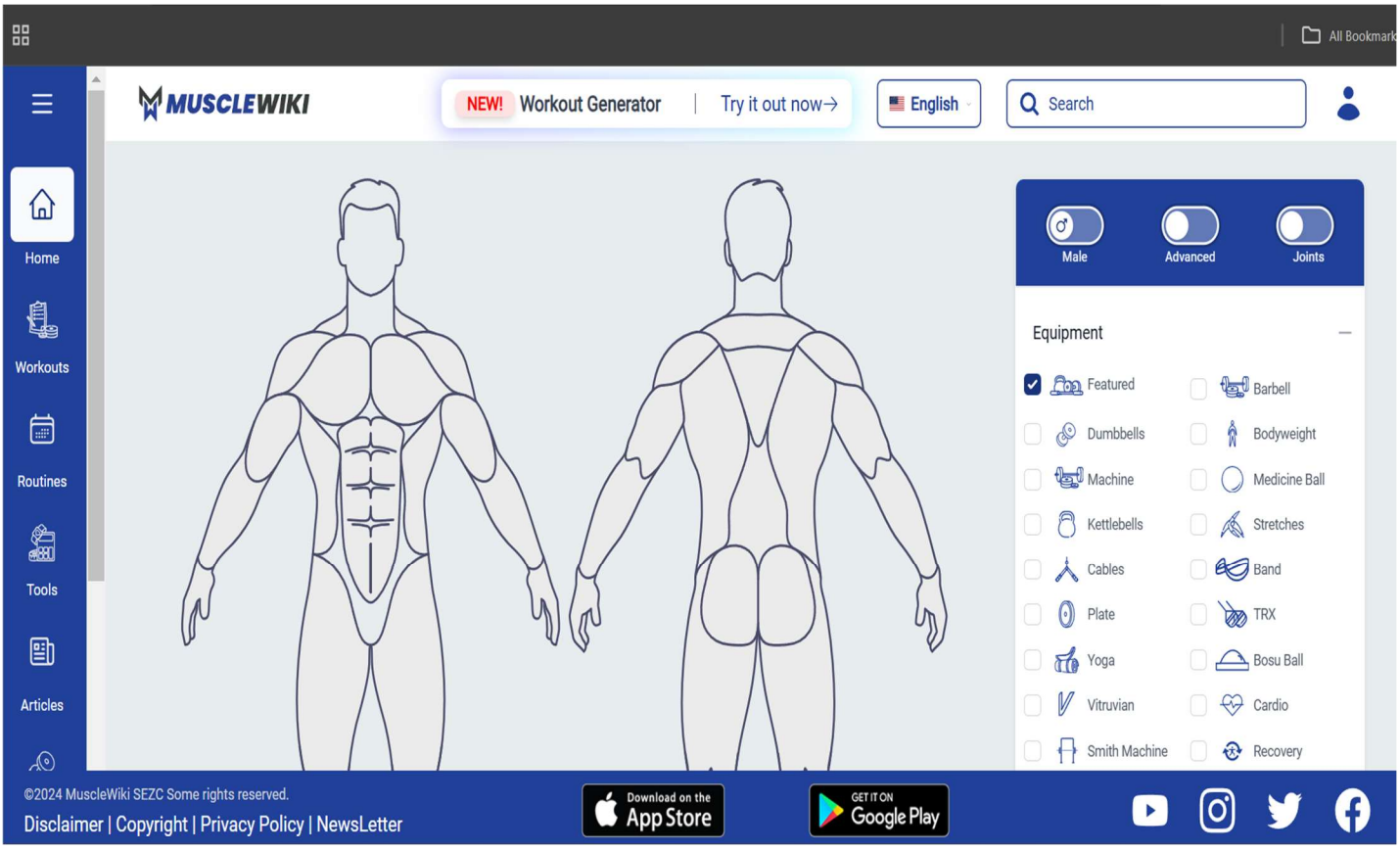


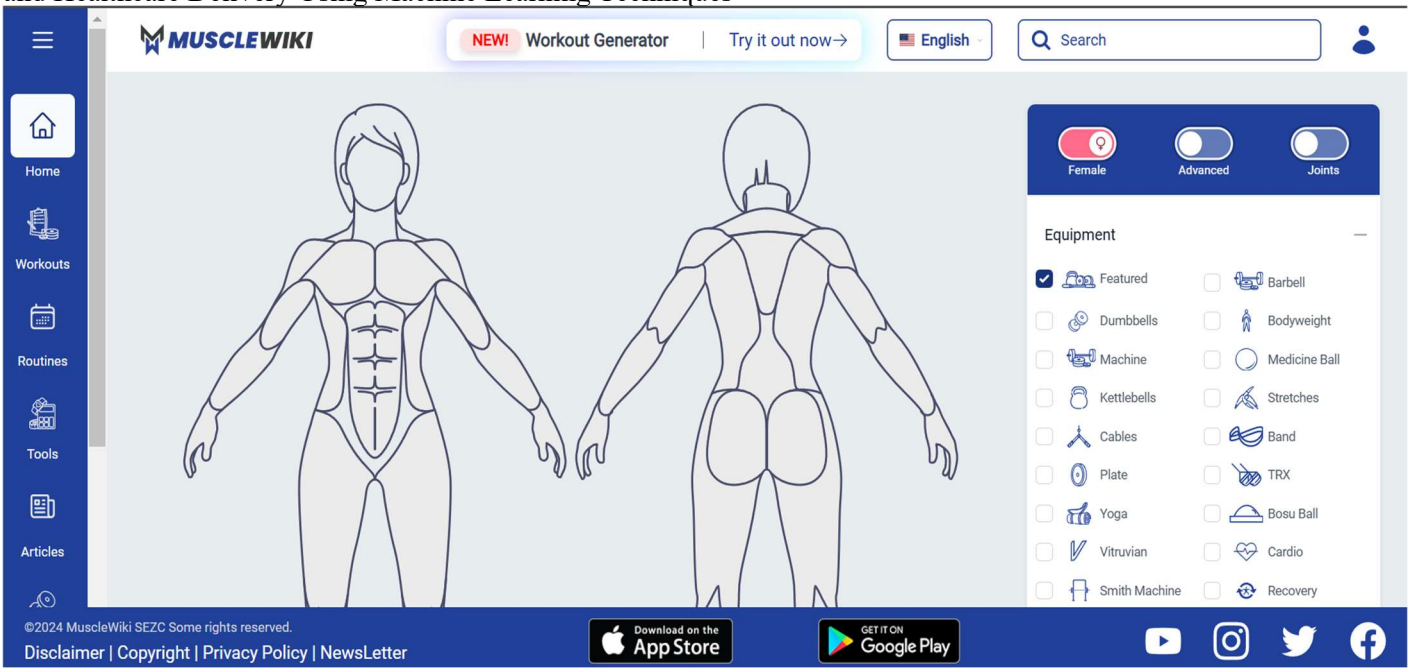


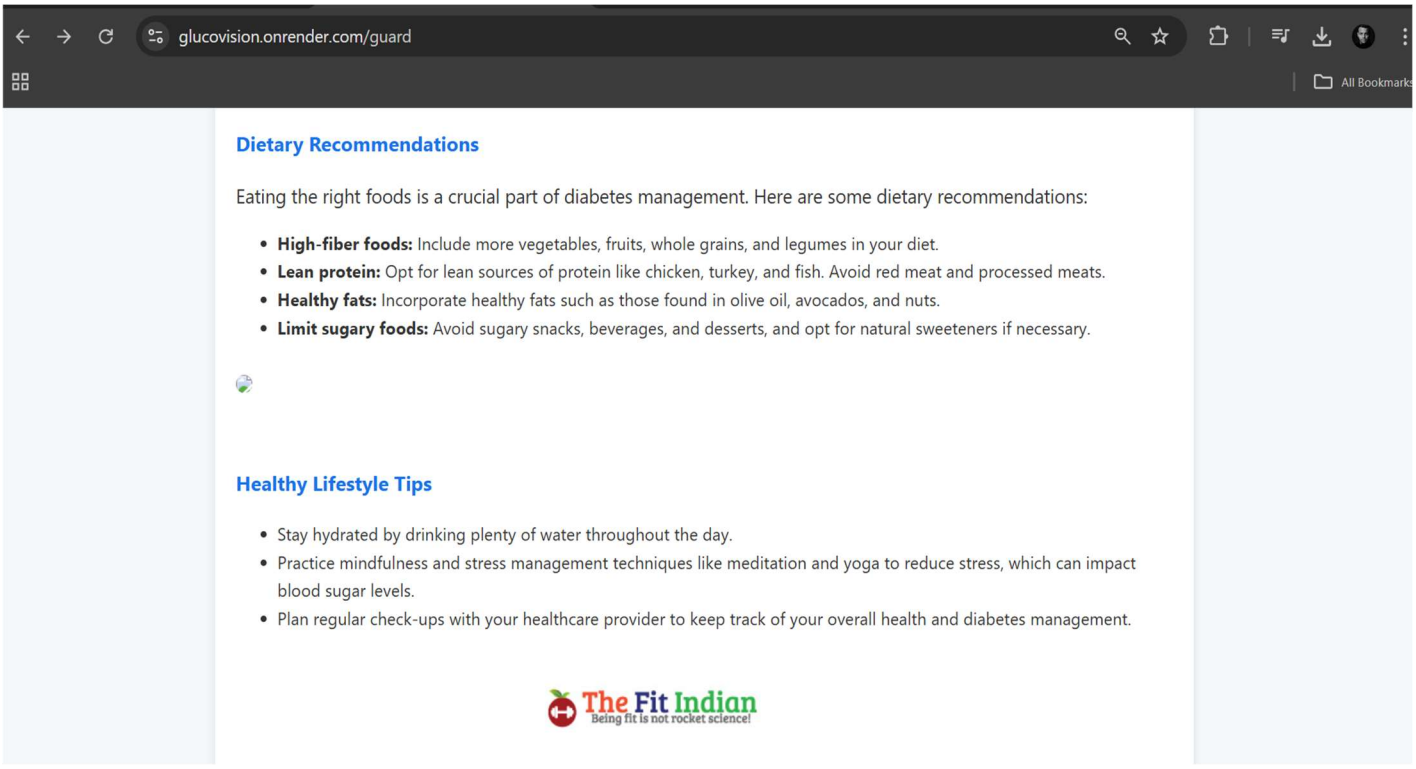
Dia Guard – A page for Health tips and Precautions



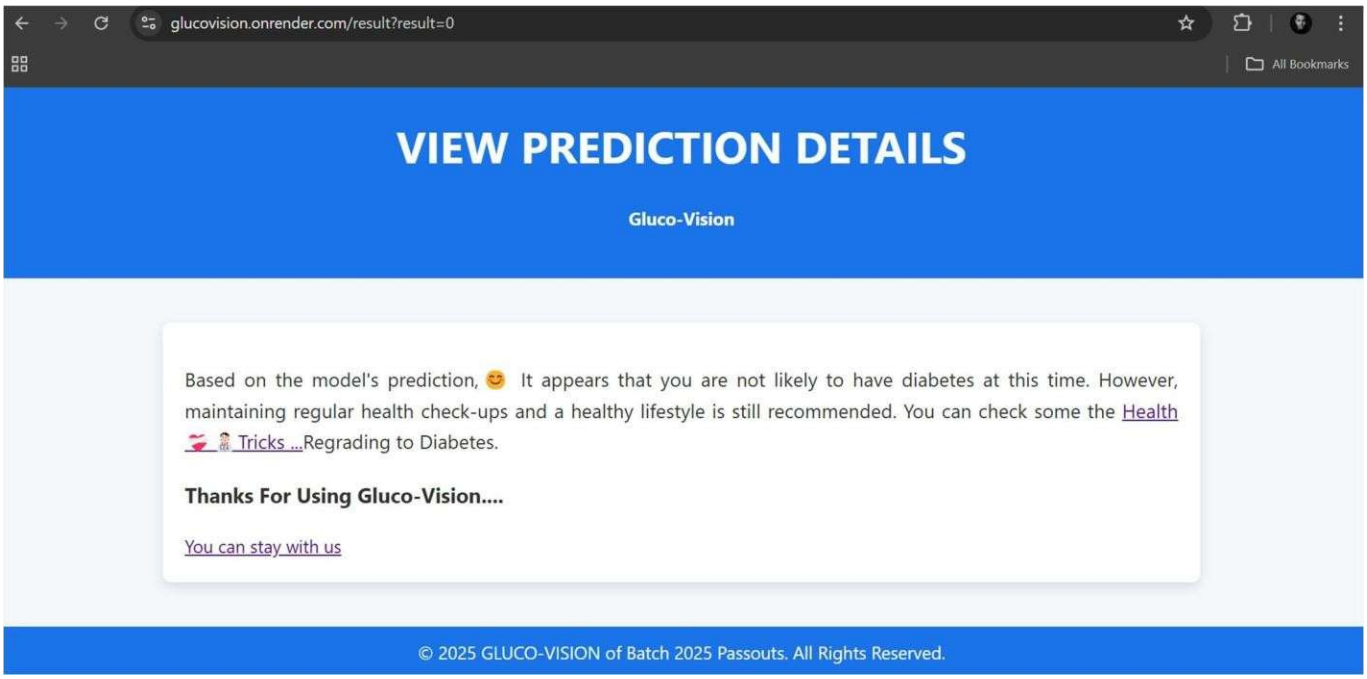

glucovision.onrender.com/guard
Preventions of Diabetes
Healthy Diet
Weight Control
Break Bad Habits
Excercise
Resources for Exercises
Dietary Recommendations

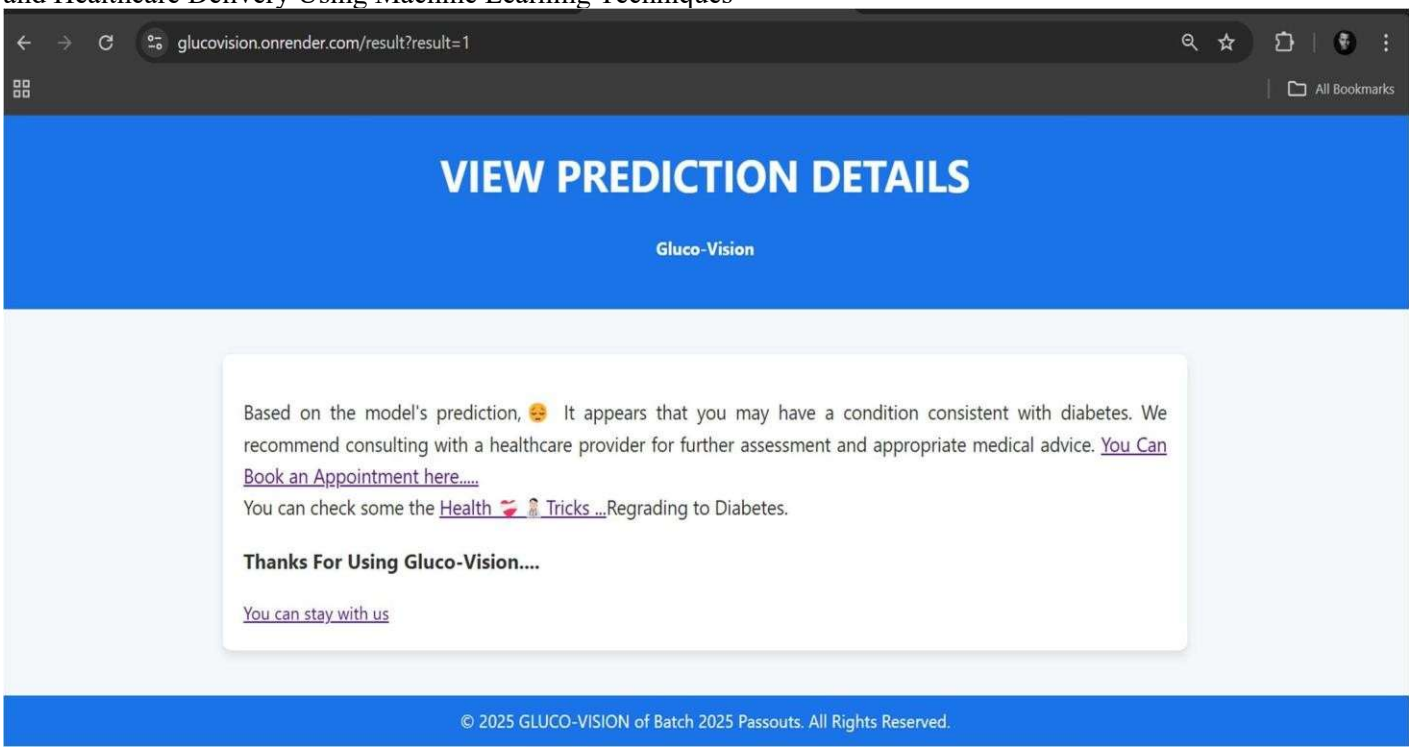

MUSCLEWIKI
NEW! Workout Generator
Try it out now→
English
Search
Home
Workouts
Routines
Tools
Articles
Male
Advanced
Joints
Equipment
Featured
Barbell
Dumbbells
Bodyweight
Machine
Medicine Ball
Kettlebells
Stretches
Cables
Band
Plate
TRX
Yoga
Bosu Ball
Vitruvian
Cardio
Smith Machine
Recovery
©2024 MuscleWiki SEZC Some rights reserved.
Disclaimer | Copyright | Privacy Policy | NewsLetter
Download on the App Store
GET IT ON Google Play


MUSCLEWIKI
NEW! Workout Generator
Try it out now→
English
Search
Home
Workouts
Routines
Tools
Articles
Female
Advanced
Joints
Equipment
Featured
Barbell
Dumbbells
Bodyweight
Machine
Medicine Ball
Kettlebells
Stretches
Cables
Band
Plate
TRX
Yoga
Bosu Ball
Vitruvian
Cardio
Smith Machine
Recovery
©2024 MuscleWiki SEZC Some rights reserved.
Disclaimer | Copyright | Privacy Policy | NewsLetter
Download on the App Store
GET IT ON Google Play

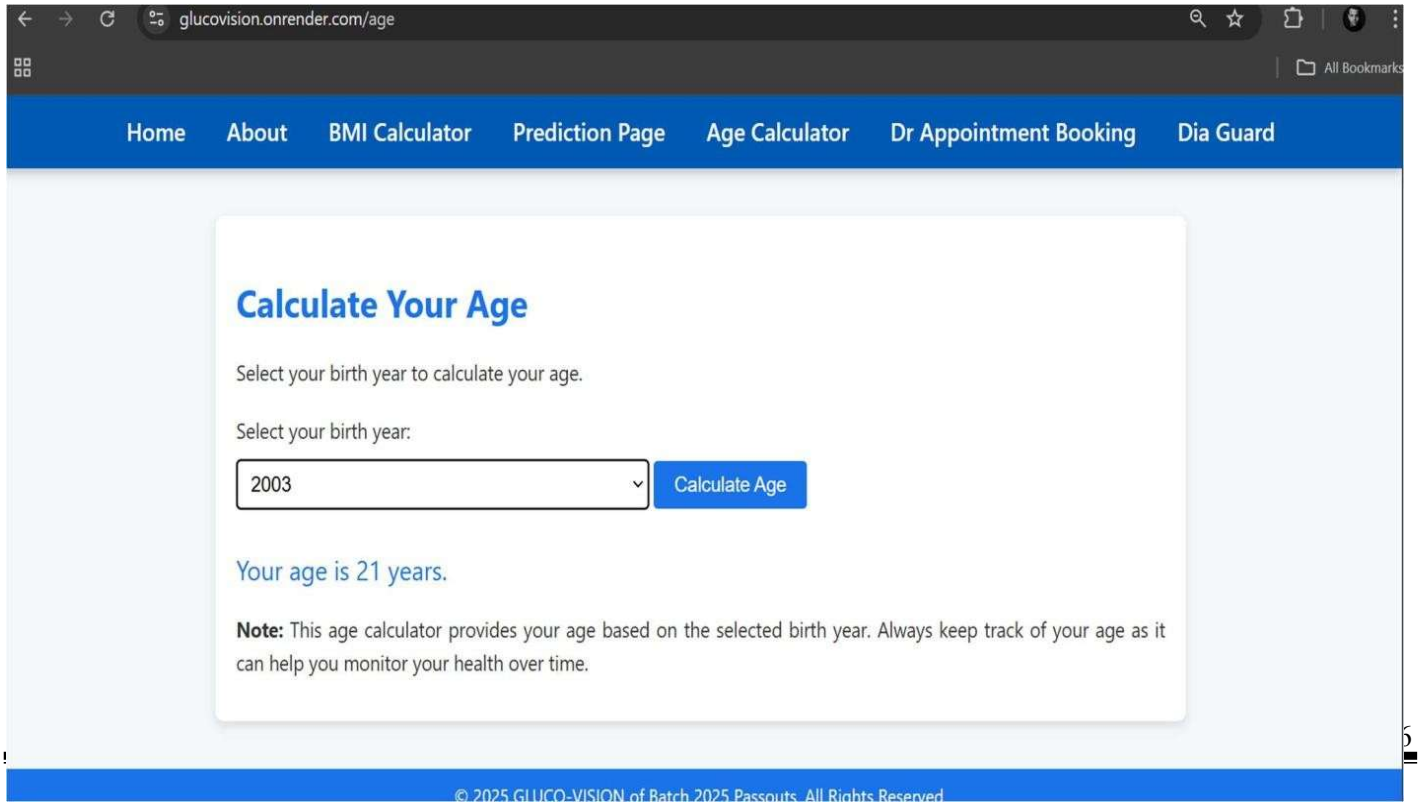

glucovision.onrender.com/guard
All Bookmarks
Dietary Recommendations
Eating the right foods is a crucial part of diabetes management. Here are some dietary recommendations:
High-fiber foods: Include more vegetables, fruits, whole grains, and legumes in your diet.
Lean protein: Opt for lean sources of protein like chicken, turkey, and fish. Avoid red meat and processed meats.
Healthy fats: Incorporate healthy fats such as those found in olive oil, avocados, and nuts.
Limit sugary foods: Avoid sugary snacks, beverages, and desserts, and opt for natural sweeteners if necessary.
Healthy Lifestyle Tips
Stay hydrated by drinking plenty of water throughout the day.
Practice mindfulness and stress management techniques like meditation and yoga to reduce stress, which can impact blood sugar levels.
Plan regular check-ups with your healthcare provider to keep track of your overall health and diabetes management.
The Fit Indian
Being fit is not rocket science!

Sample Predictions:

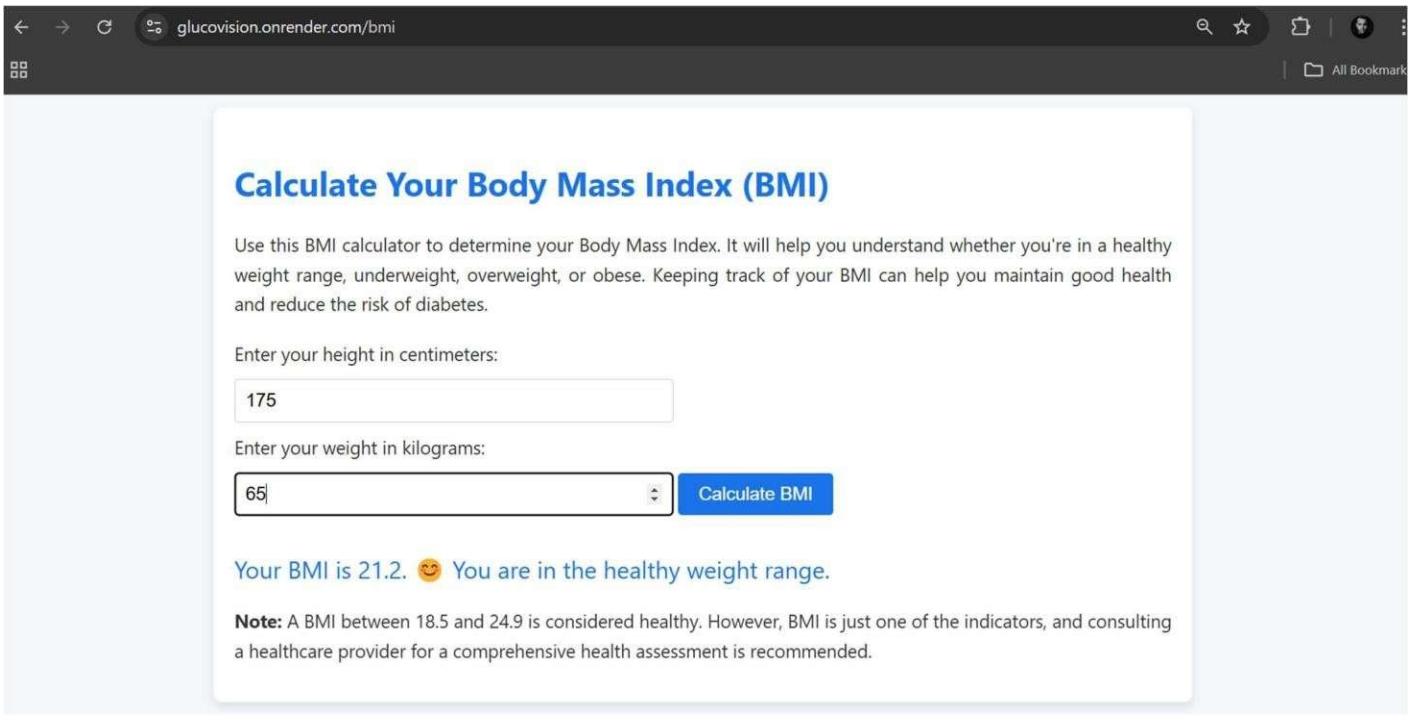




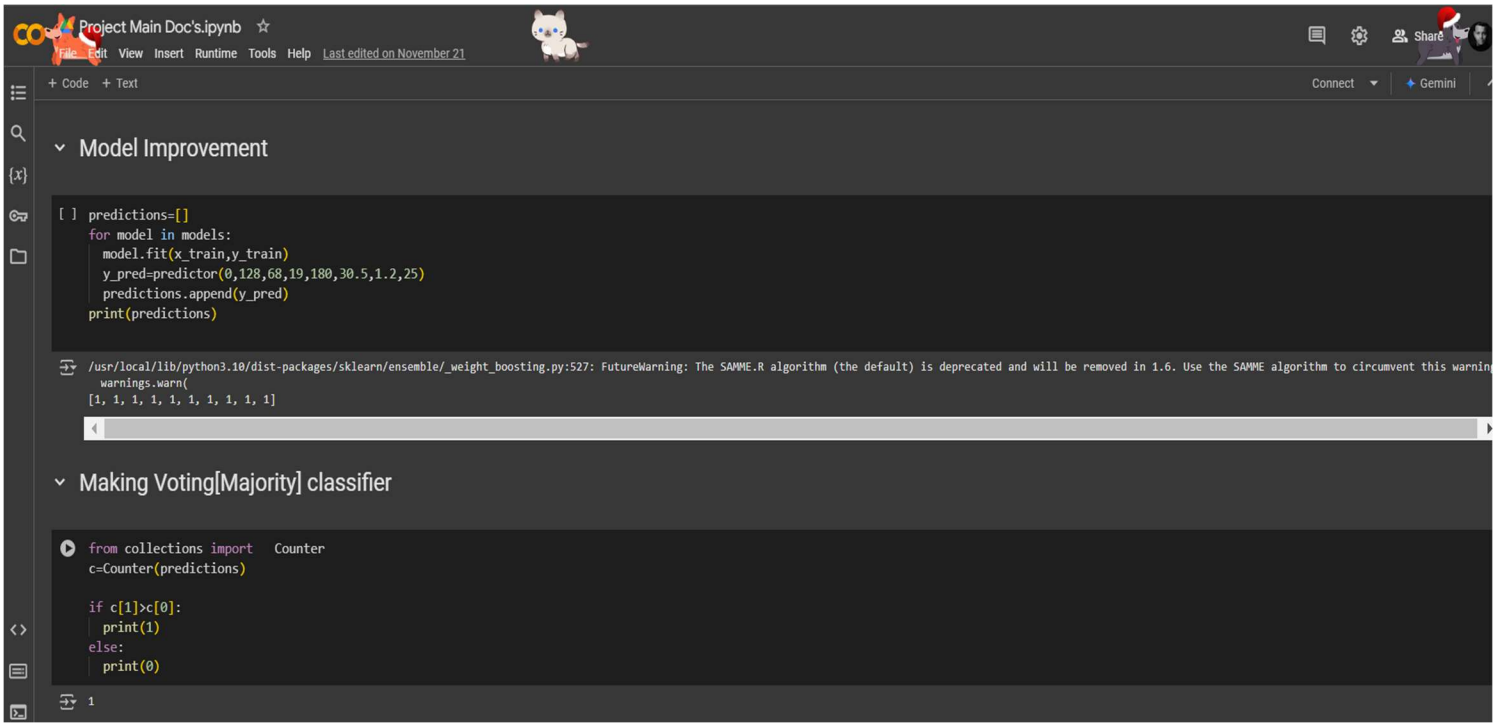
Sample Age Calculation:



Sample BMI Calculation



Proposed Method : [source code] accuracy ~=100%



After Data Feature Extraction :

Project Main Doc's.ipynb

File Edit View Insert Runtime Tools Help Last edited on November 21

+ Code + Text

Connect Gemini

Feature Extraction

```
[ ] x = df.drop('Outcome', axis=1)#Dependents
    y = df['Outcome']#Independent
```

sample records after extraction

```
[ ] x.head()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	6.0	148.0	72.0	35.0	0.0	33.6	0.627	50.0
1	1.0	85.0	66.0	29.0	0.0	26.6	0.351	31.0
2	8.0	183.0	64.0	0.0	0.0	23.3	0.672	32.0
3	1.0	89.0	66.0	23.0	94.0	28.1	0.167	21.0
4	0.0	137.0	40.0	35.0	168.0	43.1	1.200	33.0

```
y.head()
```

	Outcome
0	1
1	0
2	1
3	0

GitHub Repo Structure:

github.com/AshekCB/CSG-G03-Glucovision-Project-Doc-s

AshekCB / CSG-G03-Glucovision-Project-Doc-s

Type / to search

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

CSG-G03-Glucovision-Project-Doc-s Public

Pin Unwatch 1 Fork 0 Star 0

main 1 Branch 0 Tags Go to file t Code

AshekCB Add files via upload df3de67 · last month 6 Commits

PIP2001_-REVIEW-0_CSG-G03.pptx	Review 0	2 months ago
PIP2001_Review-1_PPT.pptx	Add files via upload	last month
Project Review 1.pdf	Add files via upload	last month
Review-1 Report 2024.docx	Add files via upload	last month
about.html	Add files via upload	2 months ago
age.html	Add files via upload	2 months ago

About

No description, website, or topics provided.

Activity
0 stars
1 watching
0 forks

Releases

No releases published
Create a new release

Git-Hub Repository Link:

https://github.com/AshekCB/CSG-G03-Glucovision-Project-Doc-s.git

Our Work is live at:

https://glucovision.onrender.com/

APPENDIX-C
ENCLOSURES

1. Journal publication/Conference Paper Presented Certificates of all students.

2. Include certificate(s) of any Achievement/Award won in any project-related event.

3. Similarity Index / Plagiarism Check report clearly showing the Percentage (%). No need for a page-wise explanation.

ABHISHEK C B

ORIGINALITY REPORT

15%	11%	8%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.scpe.org Internet Source	1%
2	jurnal.itscience.org Internet Source	1%
3	Submitted to University of Wales Swansea Student Paper	1%
4	Anshul Verma, Pradeepika Verma, Kiran Kumar Pattanaik, Lalit Garg. "Research Advances in Intelligent Computing", CRC Press, 2023 Publication	1%
5	www.e2matrix.com Internet Source	1%
6	arxiv.org Internet Source	1%
7	aiforsocialgood.ca Internet Source	1%
8	Submitted to The University of Texas at Arlington Student Paper	<1%

4. Details of mapping the project with the Sustainable Development Goals (SDG).

Our project, Gluco Vision, aligns closely with SDG-3: Good Health and Well-Being, as it leverages machine learning to provide early diabetes prediction and prevention strategies. By analyzing healthcare data, our project contributes to:

1.Improved Diagnosis and Prevention

- **Significance**: Diabetes is a major global health challenge, often undiagnosed until severe complications arise. Early detection through predictive models helps address this issue.

- **Role of Gluco Vision**: By using machine learning algorithms, our project predicts the likelihood of diabetes in individuals based on healthcare parameters like glucose levels, BMI, blood pressure, etc. This allows medical professionals to intervene early, preventing severe complications like heart disease, kidney failure, or vision problems.
- **Impact**: Early diagnosis improves treatment outcomes, reduces healthcare costs, and enhances the quality of life for individuals.

2. Healthcare Accessibility

- **Significance**: Many regions, particularly in low-resource areas, lack access to advanced diagnostic tools or specialist healthcare. Technology-driven solutions can bridge this gap.
- **Role of Gluco Vision**: By offering a web-based platform for diabetes prediction, Gluco Vision can make essential health assessments accessible to a wider population. Individuals can use the tool remotely without needing expensive lab tests or doctor consultations.
- **Impact**: Our project democratizes healthcare by providing cost-effective and scalable diagnostic solutions, ensuring underserved communities can benefit from timely health interventions.

3. Public Awareness

- **Significance**: Education about lifestyle and preventive healthcare is crucial for managing diabetes, as it is largely influenced by diet, exercise, and awareness of risk factors.
- **Role of Gluco Vision**:
    - Features like the BMI calculator and pages with precautionary advice inform users about maintaining a healthy lifestyle.
    - The project also provides an understanding of how factors like glucose, age, or family history influence diabetes risk.
    - The inclusion of solutions for managing diabetes increases users' confidence in controlling their condition.
- **Impact**: This fosters a proactive approach to health, encouraging individuals to make informed decisions and adopt healthier habits, reducing the overall burden of diabetes on society.