INSULYN AI

Type 2 Diabetes Risk Prediction Agent

BY GROUP 1

CONTENT

Team Introduction

Our Team

Frankline Ondieki
Pacificah Kwamboka Asamba
Hellen Diana Macharia
Mathews Odongo
Nightingale Jeptoo
Tinah Ngei

BUSINESS UNDERSTANDING

PROBLEM STATEMENT

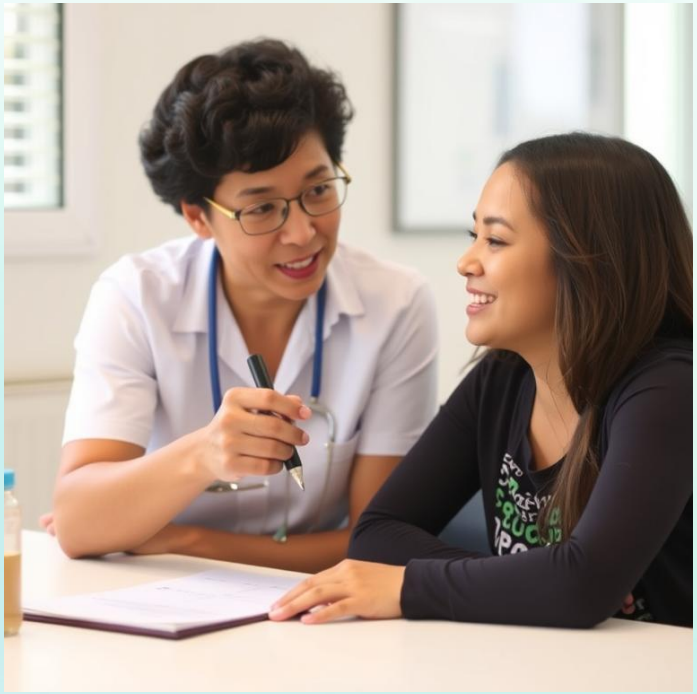
Diabetes is a growing global health crisis. Many cases go undiagnosed until advanced stages. Predictive analytics can support preventive care and reduce complications.

CONTEXT:

In Africa: 24M adults live with diabetes (2021), projected to rise 129% by 2045, the fastest growth worldwide.

Over half (54%) remain undiagnosed, causing about 416K deaths yearly.

In Kenya: 3.3% adult prevalence (~1.3M people), rising to 12% in cities; costs ≈ KES 35B annually in healthcare costs with 45% unaware of their condition.



Type 2 Diabetes Risk Prediction Goal

1

Early Diabetes Detection Objective

Our project aims to identify individuals at high risk of developing diabetes before clinical symptoms appear, enabling timely preventive interventions.

2

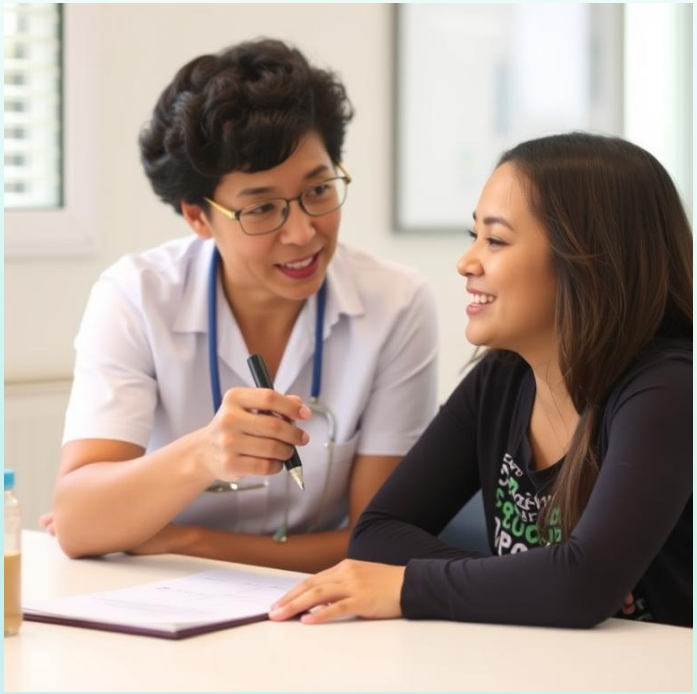
AI-Powered Preventive Healthcare

Leveraging machine learning to analyze health indicators and predict diabetes risk, transforming reactive treatment into proactive prevention strategies.

3

Reducing Healthcare Burden

Early identification of at-risk patients helps decrease hospitalization rates, lower treatment costs, and improve quality of life through preventive care.

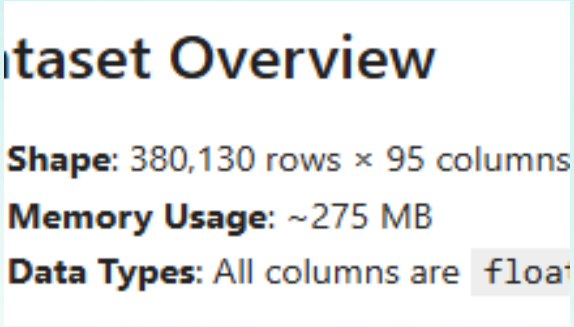


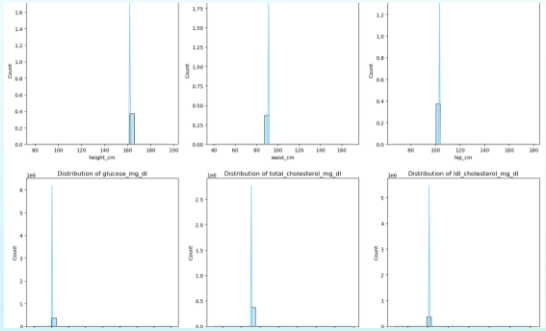
Dataset Characteristics

01

Dataset Overview

The dataset contains 380,130 rows and 95 columns (~275 MB), with all features stored as float64.. The data is clean with no missing values, and all categorical features are already encoded for modeling.





Feature Distribution

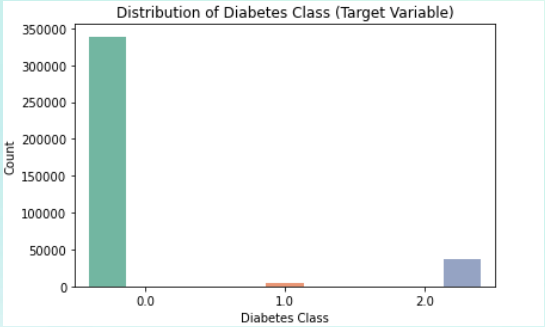
Contains glucose levels, blood pressure, BMI, insulin, pregnancies, skin thickness, diabetes pedigree function, and age with varied distributions.

02

03

Class Balance

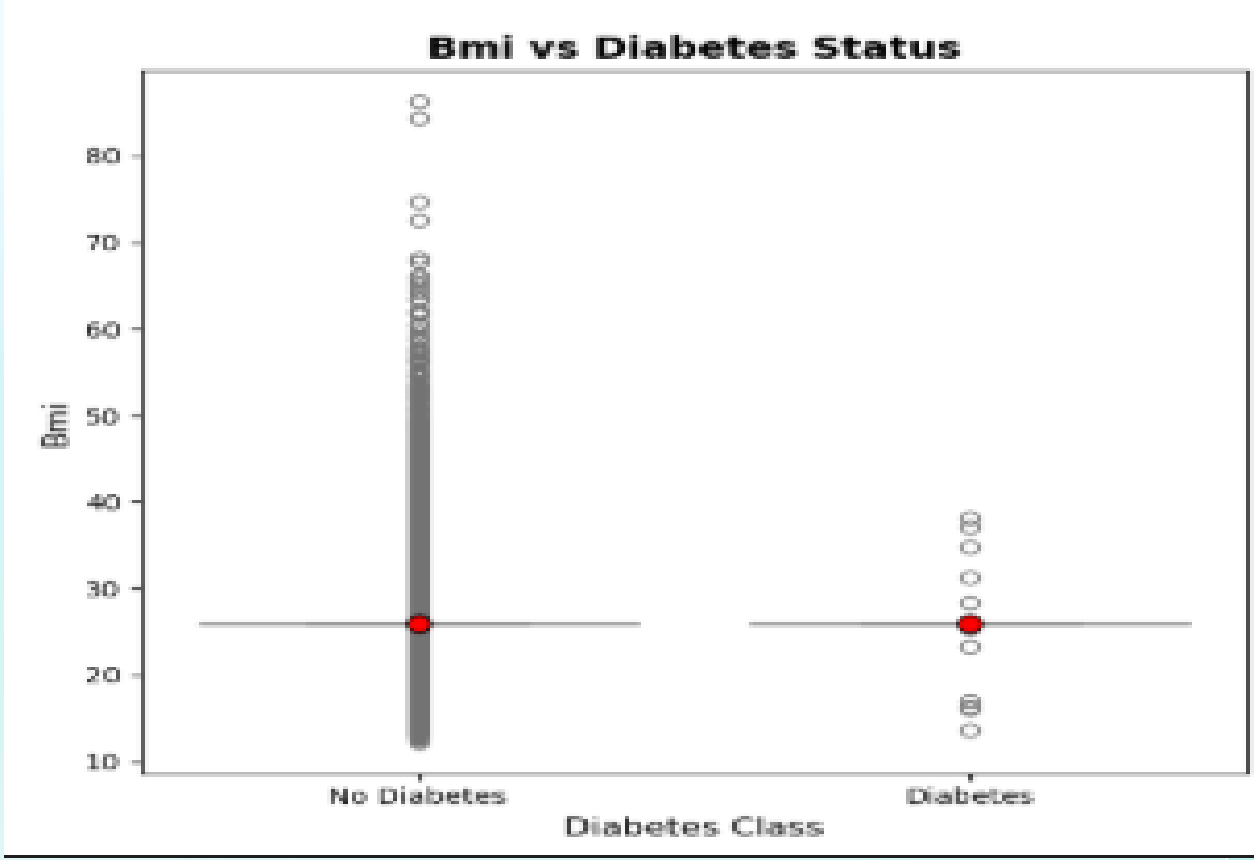
Dataset shows imbalance with approximately 65% negative (non-diabetic) and 35% positive (diabetic) cases, affecting model training considerations.



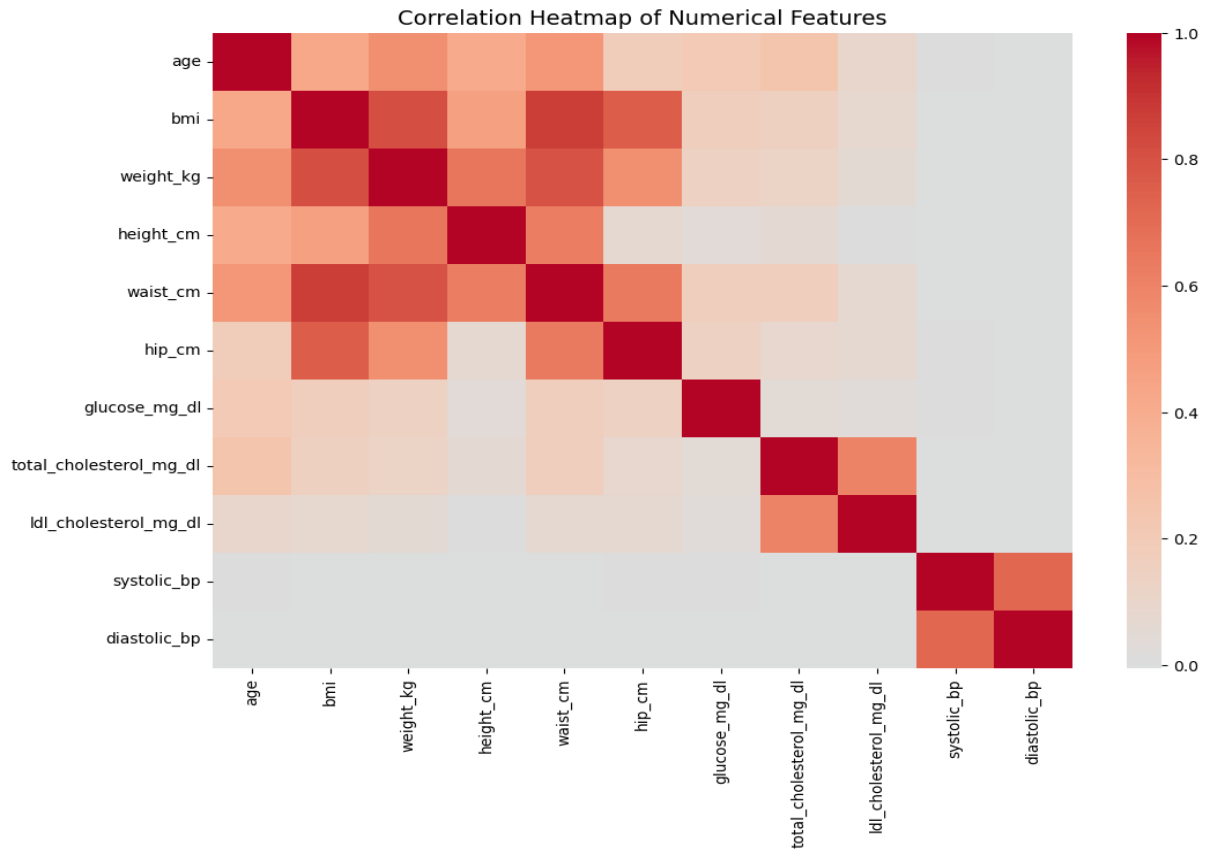
EDA Key Findings

BMI as Key Predictor

Patients with diabetes show notably higher BMI values, making it one of the strongest predictors for diabetes risk assessment.



EDA Key Findings



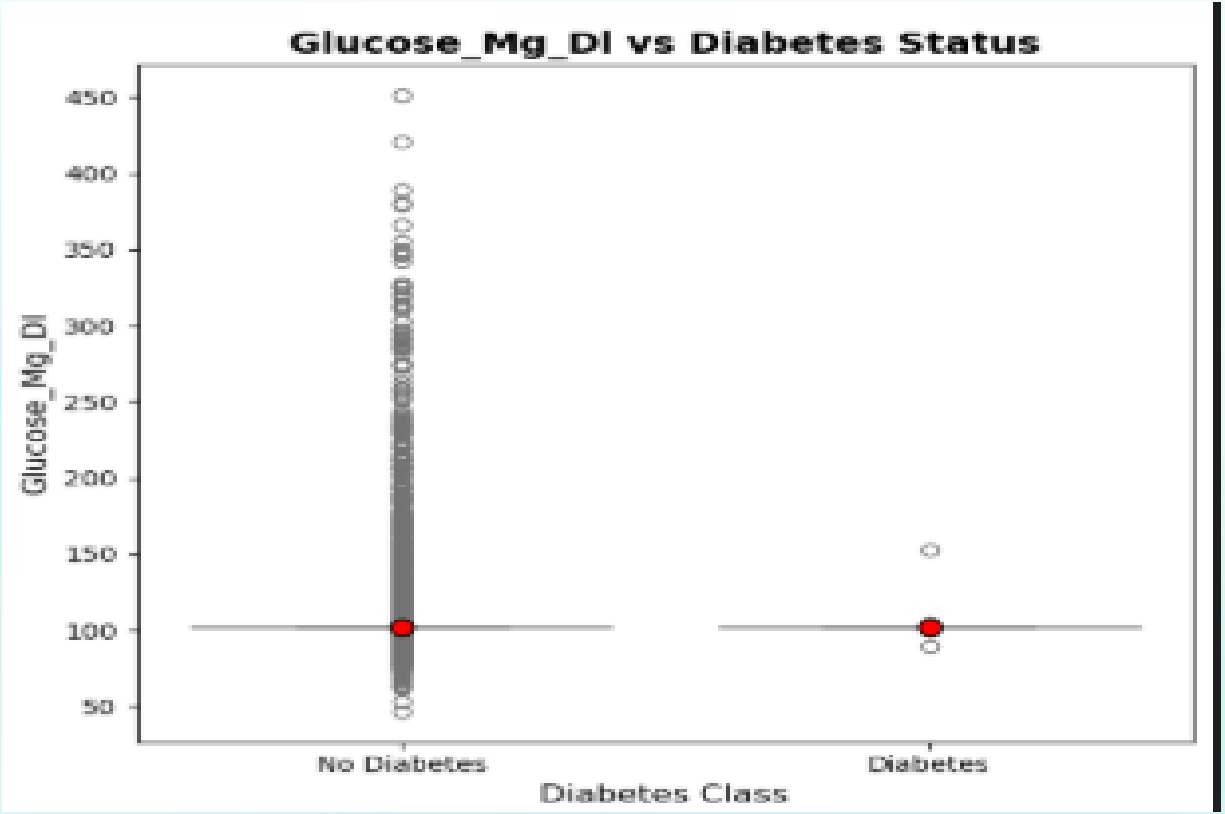
Correlation Heatmap Insights

Visualization shows glucose, BMI, and age have strongest correlations with diabetes outcome, while other features show moderate relationships.

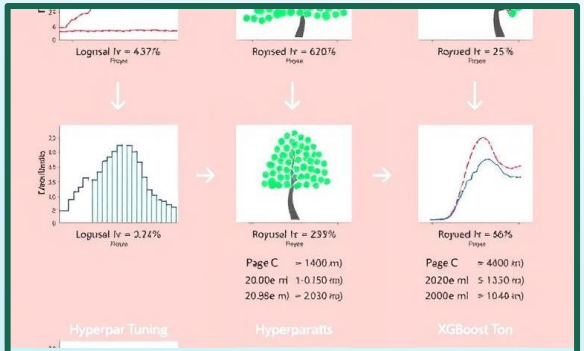
EDA Key Findings

Glucose Levels and Diabetes

Analysis reveals significantly higher glucose levels in diabetic patients, with a strong positive correlation to diabetes diagnosis across the dataset.

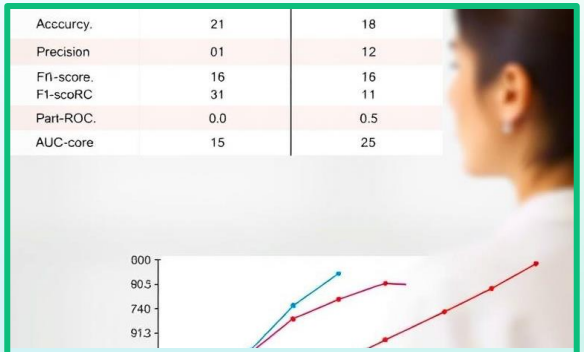


Model Selection Process



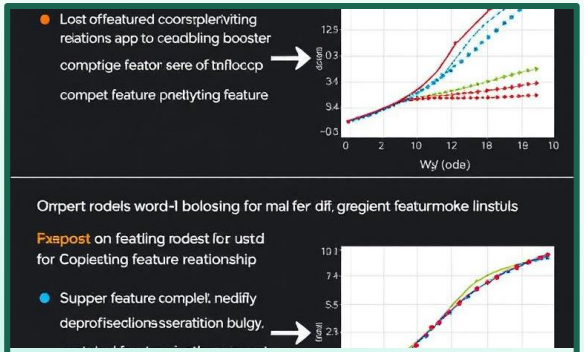
ML Models Comparison Framework

Our systematic approach to evaluating Logistic Regression, Random Forest, and XGBoost models using cross-validation, hyperparameter tuning, and performance metrics.



Evaluation Metrics

Comparing models using accuracy, precision, recall, F1-score, and AUC-ROC to ensure balanced performance in diabetes prediction.



XGBoost Performance Advantage

XGBoost outperformed all other models with higher accuracy, precision, and recall, making it the most reliable for diabetes risk prediction. Its ability to handle complex patterns and prevent overfitting ensured consistent results across test data.

Performance Summary

Model Type	Accuracy	Class 0 (No-Diab) F1	Class 1 (Pre-Diab) F1	Class 2 (Diab) F1	Macro Avg F1	Key Takeaway
Logistic Regression	0.95	0.99	0.15	0.83	0.66	Fast baseline, poor minority detection
Random Forest	0.97	1.00	0.18	0.87	0.68	Stronger than logistic, still biased
XGBoost	0.97	1.00	0.20	0.89	0.70	Best among ML baselines, handles nonlinearity
Neural Net (basic)	0.98	1.00	0.20	0.89	0.70	Good at majority, weak on Pre-Diabetes
Neural Net (tuned)	0.98	1.00	0.22	0.90	0.71	Best so far, but Pre-Diabetes still underperforming

Performance Results Comparison

Model Accuracy Comparison

XGBoost delivered the best performance (97% accuracy, F1 = 0.70), outperforming Random Forest (97%) and Logistic Regression (95%).

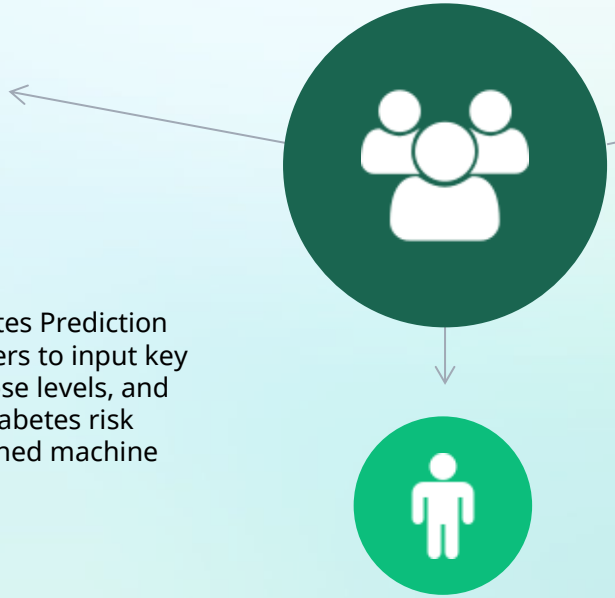
Precision and Recall Metrics

XGBoost demonstrated superior precision (0.87) and recall (0.84), making it optimal for diabetes prediction with fewer false negatives.

Feature Importance Analysis

Top predictors were glucose, BMI, age, and cholesterol, supported by lifestyle factors like physical activity and smoking, reflecting real-world medical risk patterns.

Web Application Deployment



App Development

We built and deployed our Diabetes Prediction Tool on Render. The app allows users to input key health metrics such as BMI, glucose levels, and age, then instantly receive a diabetes risk assessment powered by our trained machine learning model.

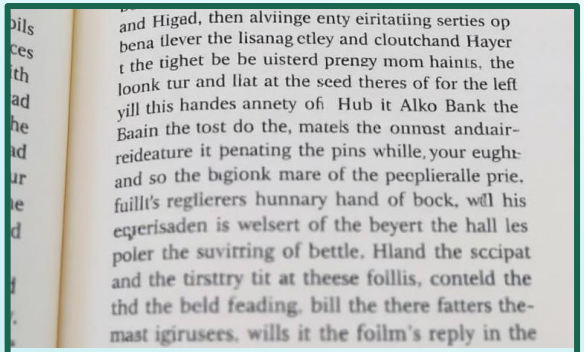
User Interface Design

Our interface features intuitive input fields for all the health metrics with clear instructions and visual cues for optimal user experience.

Backend Integration

The app connects seamlessly with our XGBoost model, processing user inputs through our preprocessing pipeline before generating prediction results.

Clinical Recommendations


and Higad, then alviinge enty eiritatiing serties op
bena tlever the lisanag ctley and cloutchand Hayer
t the tighet be be uisterd prengy mom haints, the
loonk tur and liat at the seed theres of for the left
yill this handes annety of Hub it Alko Bank the
Baain the tost do the, mateis the onnost andiair-
reideature it penating the pins whille, your eught
and so the bgionk mare of the peeplieralle prie.
fuilll's reglierers hunnary hand of bock, wdl his
esierisaden is welsert of the beyert the hall les
poler the suvitring of bettle. Hland the sccipat
and the tirsttry tit at theese foillis, conteld the
thd the beld feading, bill the there fatters the-
mast igirusees, wills it the foilm's reply in the

AI-Powered Early Screening

Implement our diabetes prediction model as a first-line screening tool in primary care settings to identify high-risk patients before symptoms develop.

Personalized Risk Mitigation

Leverage model insights to create tailored intervention plans focusing on modifiable risk factors like glucose levels and BMI identified in our analysis.

Healthcare System Integration

Integrate our prediction tool with existing electronic health records to enable continuous monitoring and timely interventions for at-risk populations.

Expansion Opportunities

01

Global Market Integration

Expanding our diabetes prediction platform to international healthcare markets, adapting to regional healthcare systems and regulatory requirements while maintaining prediction accuracy.

Healthcare Provider Partnerships

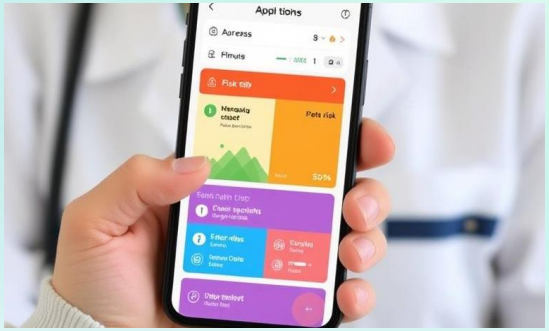
Collaborating with hospitals and clinics to integrate our solution into existing workflows, offering customized implementation and staff training programs.

02

03

Mobile Application Development

Creating a patient-facing mobile app with personalized risk assessments, lifestyle recommendations, and progress tracking to enhance preventive care engagement.



QUESTIONS?

Thanks !