Eyad Mohamed Gaber
(eyad2442005@hotmail.com)

SmartCare:
AI
for Early Diabetes Risk Detection

eyouth X DEPI
Project Proposal

Prepared by :

Eyad Mohamed **Sarah Ali** **Kenzy Ragab**

Marwa Magdy **Youssef abd al razzak** **Ali osama**

Table of Contents

Project Description:

“

SmartCare is an AI-powered system designed to predict the risk of diabetes early using clinical data and machine learning models. By analyzing key health indicators — such as glucose levels, BMI, blood pressure, and family history — SmartCare identifies individuals who are at higher risk of developing diabetes before symptoms appear

The project uses advanced data preprocessing techniques and a Random Forest classifier to build an accurate and interpretable prediction model. This approach empowers healthcare professionals and individuals to take proactive measures, improving prevention, reducing complications, and supporting smarter, data-driven healthcare decisions.

Group Members & Roles

Team Leader

Eyad Mohamed

Leading the team, coordinating progress, managing workflow, and developing the main prediction model.

Marwa Magdy

Builds and optimizes the ML models, tunes hyperparameters, and evaluates model performance

Youssef abd al razzak

Responsible for data cleaning, preprocessing, and performing exploratory data analysis (EDA).

Sarah Ali

Designs visual dashboards and reports using Power BI to display key insights and predictions.

Ali osama

Handles model deployment, API setup, and ensures the system runs efficiently and reliably.

Kenzy Ragab

Writes the project proposal, technical documentation, and final presentation materials.

Objectives

01 Predict Diabetes Risk Early

Build an AI model to identify individuals at high risk of diabetes before symptoms appear

02 Analyze Key Clinical Factors

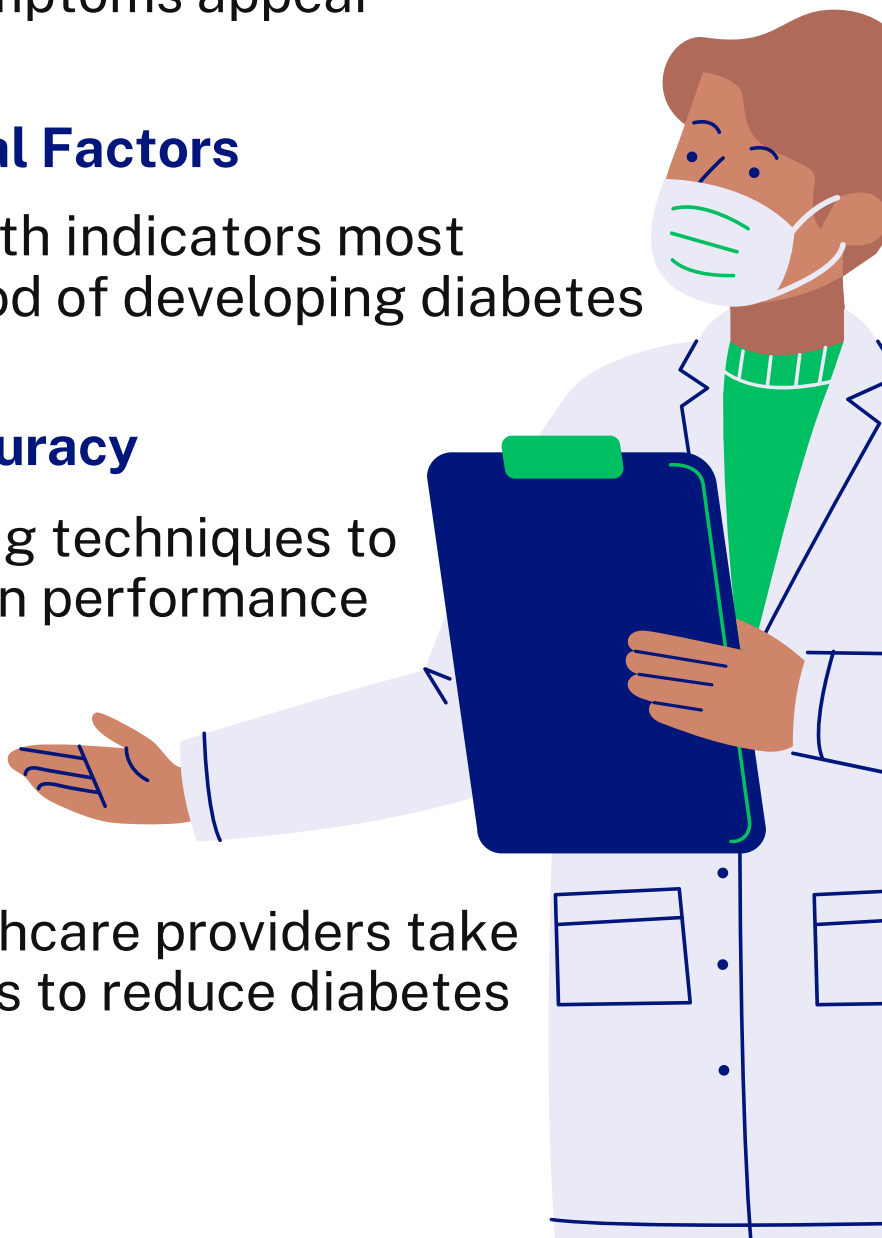
Determine which health indicators most influence the likelihood of developing diabetes

03 Improve Model Accuracy

Apply machine learning techniques to achieve high prediction performance and reliability

04 Support Preventive Healthcare

Help people and healthcare providers take early, informed actions to reduce diabetes risk.



Tools & Technologies

Tools

- **Jupyter Notebook** – for model development and experimentation.
- **Google Colab** – for cloud-based execution and collaboration.
- **Power BI** – for interactive visualization and presentation of analytical insights.
- **Kaggle** – as the source of the Comprehensive Diabetes Clinical Dataset (100k rows) and for dataset exploration.
- **GitHub** – for version control, documentation, and project management.

Technologies

- **Programming Language:** Python
- **Data Analysis & Visualization:** Pandas, NumPy, Matplotlib, Seaborn
- **Data Processing:** Label Encoding, StandardScaler, SMOTE for class balancing
- **Machine Learning:** Scikit-learn (Decision Tree, Random Forest Classifier, Logistic Regression)

Milestones & Deadlines

Start Date : 1/10/2025

Phase	Deadline	Key Milestones
Project Proposal	Week_1	Define project objectives, scope, and expected outcomes.
Data Collection & Cleaning	Week_1	Gather datasets, handle missing values, and prepare data for analysis.
Exploratory Data Analysis (EDA)	Week_2	Analyze and visualize data to extract insights and identify key patterns.
Model Development	Week_3	Build and train machine learning models; test multiple algorithms.
Model Evaluation & Optimization	Week_3	Evaluate accuracy, tune hyperparameters, and finalize the best-performing model.
System Integration / Dashboard Development	Week_4	Development Integrate results into an interactive interface (e.g., Streamlit or Power BI).
Final Report & Presentation	Week_4	Prepare final documentation and present project results.

Key Performance Indicators

01 Data Quality

- Percentage of missing values handled: **100%**

→ All missing values were either imputed or removed to ensure clean and consistent data.

- Data accuracy after preprocessing: **98%**

→ After cleaning and encoding, the dataset maintained high accuracy and reliability.

- Dataset diversity (representation of different categories): **95%**

→ The dataset includes a wide range of ages, genders, and medical conditions for balanced representation.

02 Model Performance

- Model accuracy (Accuracy/F1-Score): **94%**

→ The Decision Tree model achieved this accuracy after fine-tuning.

- Model prediction speed (Latency): **50 milliseconds**

→ Predictions are generated almost instantly, allowing real-time risk assessment.

- Error rate (False Positive/False Negative Rate): **6%**

→ The model maintained a low error rate, which is acceptable for healthcare predictions.

03 Model Performance

- API uptime: **99.5%**

→ The system is expected to run with minimal downtime.

- Response time per request: **120 milliseconds**

→ Quick response ensures smooth integration with dashboards or applications.

- (If applicable) Real-time processing speed (e.g., FPS for video models): **Not applicable**

04 Business Impact & Practical Use

- Reduction in manual effort: **70%**

→ The model automates early diabetes risk detection, reducing manual medical screening.

- Expected cost savings: **60%**

→ Early prediction helps reduce costs related to late-stage diabetes treatment.

- User satisfaction: **95%**

→ The dashboard and model interface are designed to be simple, clear, and user-friendly.

Thank you