

# Project Proposal



**Proposed By :**

Habiba Ashrsf Seif  
Ali Osama Ali  
Magdy Mohamed Hassan

# Diabetes Prediction Using Machine Learning

- **Idea**

The main idea of this project is to develop a Machine Learning model capable of predicting whether a person is likely to have diabetes based on specific medical features such as age, BMI (Body Mass Index), blood pressure, glucose level, number of pregnancies, and insulin levels. The goal is to assist doctors and patients in early detection, allowing timely medical intervention and reducing the risk of severe health complications.

- **Introduction**

Diabetes Mellitus is one of the most widespread chronic diseases globally, affecting millions of people every year. Early diagnosis plays a crucial role in preventing life-threatening complications such as cardiovascular diseases, kidney failure, and nerve damage.

With the rapid advancement of Artificial Intelligence (AI) and Machine Learning (ML), these technologies have become powerful tools in medical data analysis and disease prediction.

This project aims to leverage ML algorithms to predict diabetes based on medical parameters, providing a smart and data-driven approach to early diagnosis and preventive healthcare.

## • Problem Statement

Traditional methods for diagnosing diabetes rely heavily on laboratory tests and clinical evaluations, which can be time-consuming and may delay early intervention.

Moreover, many individuals remain undiagnosed due to mild or hidden symptoms in the early stages.

Therefore, there is a strong need for an intelligent prediction system that can analyze medical data and predict diabetes risk quickly and accurately, helping healthcare providers make informed decisions.

## • **Objectives**

- To develop and train a Machine Learning model for predicting diabetes.
- To identify the most significant medical features influencing diabetes risk.
- To compare the performance of multiple ML algorithms such as Logistic Regression, Decision Tree, and Random Forest.
- To improve model accuracy using data preprocessing and cross-validation techniques.
- (Optional) To build a simple user interface that allows users to input their data and instantly receive prediction results.

## • **Related Works**

- Pima Indians Diabetes Dataset (UCL Repository):
  - A widely used dataset containing medical records of female patients from the Pima Indian community, commonly used for diabetes prediction research.
- Sisodia et al. (2018):
  - Applied Logistic Regression and Random Forest models to predict diabetes with an accuracy of around 85%.
- Kavakiotis et al. (2017):
  - Provided a comprehensive review of AI and ML applications in diabetes diagnosis and complication prediction.
- Kaggle Projects:
  - Multiple open-source projects on Kaggle demonstrate the use of SVM, KNN, and ANN for classifying patients as “Diabetic” or “Non-Diabetic.”