AI BASED DIABETES PREDICTION SYSTEM

PROJECT TITLE	AI based diabetes prediction system
SKILLS TAKEN AWAY	Python script EDA UI deployment
DOMAIN	Medical

PROBLEM DEFINITION:

The problem is to build an AI-powered diabetes prediction system that uses machine learning algorithms to analyse medical data and predict the likelihood of an individual developing diabetes. The system aims to provide early risk assessment and personalized preventive measures, allowing individuals to take proactive actions to manage their health.

APPROACH:

1.DATA COLLECTION:

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective is to predict based on diagnostic measurements whether a patient has diabetes. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

Pregnancies: Number of times pregnant

Glucose: Plasma glucose concentration a 2 hours in an oral glucose tolerance test

Blood Pressure: Diastolic blood pressure (mm Hg)

Skin Thickness: Triceps skin fold thickness (mm)

Insulin: 2-Hour serum insulin (mu U/ml)

BMI: Body mass index (weight in kg/(height in m)^2)

Diabetes Pedigree Function: Diabetes pedigree function

Age: Age (years)

Outcome: Class variable (0 or 1)

DATASET LINK: https://www.kaggle.com/datasets/mathchi/diabetes-data-set

2.DATA PREPROCESSING:

Clean the data to handle missing values, outliers, and inconsistencies.
Normalize or standardize the data to ensure that features have similar scales.
Encode categorical data (e.g., gender) into numerical values

3.EXPLORATORY DATA ANALYSIS:

- **Understand the Data:** EDA helps to get a grasp of the dataset's structure, size, and content. This includes identifying the types of variables (categorical, numerical, etc.) and their relationships.
- **Detect Patterns and Anomalies:** By visualizing data through plots, charts, and statistical summaries, we can identify patterns, trends, and potential outliers or anomalies in the data.
- **Formulate Hypotheses:** EDA often leads to the generation of hypotheses about the data, which can guide further analysis and modelling.

4.FEATURE ENGINEERING:

- **Creating New Features:** we might generate new features that could capture important information about diabetes risk.
- **Encoding Categorical Variables:** The dataset contains categorical variables like gender, we may convert them into numerical representations through techniques like one-hot encoding or label

encoding. This allows machine learning models to work with categorical data

- **Feature Scaling:** Depending on the machine learning algorithms you're using; you might standardize or normalize numerical features to ensure they have similar scales. This can help models converge faster and perform better.

5.MODEL BUILDING AND EVALUATION METRICES:

Use of pickle module to dump and load models. And these machine learning models are evaluated based on their performance using metrics like accuracy, precision, recall, F1-score, and ROC-AUC.

6.MODEL GUI:

A "model GUI" refers to a graphical user interface (GUI) is designed to interact with or visualize machine learning models. These interfaces are created to make it easier for users, often non-technical ones, to interact with and understand the outputs of machine learning models.

USE CASES:

- **Early Diabetes Risk Assessment:**
  Individuals can receive early risk assessments to determine their likelihood of developing diabetes. This empowers them to take proactive measures to manage their health.
- **Personalized Preventive Measures:**

  Individuals can receive personalized recommendations for preventive measures based on their risk assessments. These recommendations may

include lifestyle changes, dietary modifications, exercise routines, and regular check-ups.

- **Improved Health Management**:

  Those at higher risk of diabetes can work closely with healthcare professionals to develop tailored health management plans, potentially reducing the progression of the disease.

- **Reduction in Complications:**

  Early intervention and preventive measures can help reduce the risk of diabetes-related complications, such as cardiovascular problems, neuropathy, and kidney disease.

OUTCOMES:

Gain valuable skills in data analysis, machine learning, data preprocessing, and model building, GUI, Libraries (NumPy, pandas, Matplotlib) which are highly sought-after in the field of data science and artificial intelligence.
GAIN experience in EDA visualisation
Learn and apply ML Technology
Built optimized ML models
Develop web applications to show case ML model and predict and new data
Understand the challenge and best practices in the medical and how ML helps to solve them