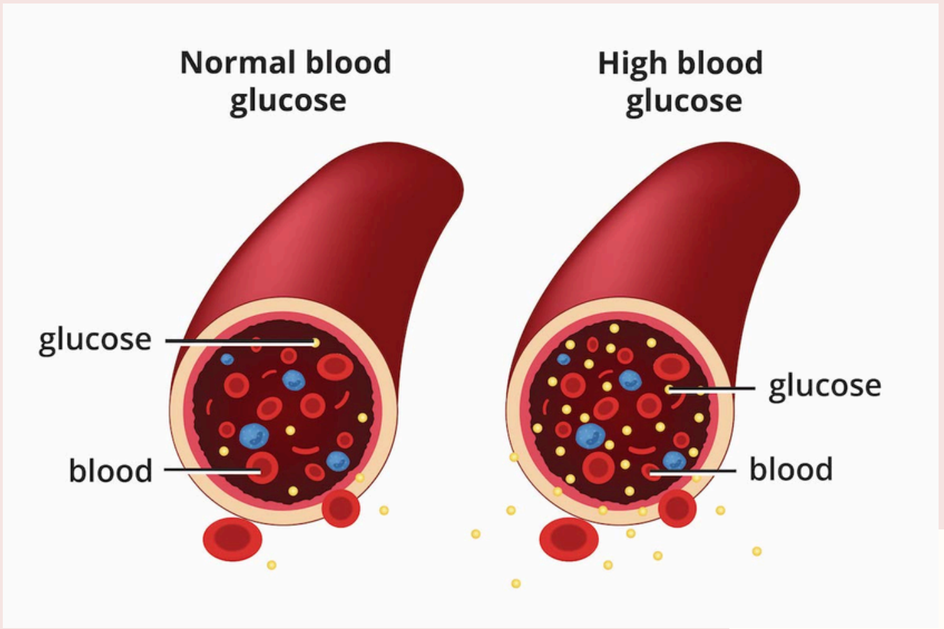
Diabetes Early Detection



Features of Dataset:

The Diabetes Early Detection Dataset contains 35 features (columns), each representing either a patient attribute, medical measurement, or symptom. Here's a breakdown of the different features:

1. Patient Demographics

- **Patient ID**: Unique identifier for each patient.
- **Name**: Patient's name (likely anonymized).
- **Age:** Age of the patient.
- **Gender:** Male/Female.

2. **Medical Measurements**

- **HbA1c_level**: Glycated hemoglobin level, used to assess blood sugar over 2-3 months.
- **BloodPressure:** Blood pressure (possibly systolic).
- **Blood_glucose_level**: Blood glucose concentration.
- **BMI:** Body Mass Index, a measure of body fat based on height and weight.
- **Urea Test**: Blood urea level.
- **Cr Test**: Creatinine level, indicating kidney function.
- **Chol Test**: Cholesterol level.
- **TG Test**: Triglyceride level.
- **HDL Test**: High-density lipoprotein (good cholesterol).
- **LDL Test**: Low-density lipoprotein (bad cholesterol).
- **VLDL Test**: Very low-density lipoprotein.

3. Health Conditions

- **Diabetes**: Indicates if the patient has previously been diagnosed with diabetes (Normal/Abnormal).
- **Hypertension**: High blood pressure (Yes/No).
- **Heart_disease**: Presence of heart disease (Yes/No).
- **Smoking_history**: Whether the patient has a history of smoking (Yes/No).
- **Skin Thickness**: Skin fold thickness, often used as a proxy for body fat.

4. Symptoms and Risk Indicators

- **Polyuria:** Excessive urination.
- **Polydipsia**: Excessive thirst.
- **Sudden weight loss**
- **Weakness**
- **Polyphagia:** Excessive hunger.
- **Genital thrush:** Fungal infection.
- **Visual blurring**
- **Itching**
- **Irritability**
- **Delayed healing**: Wounds healing slowly.
- **Partial paresis:** Muscle weakness or partial paralysis.

- **Muscle stiffness**
- **Alopecia**: Hair loss.
- **Obesity:** Presence of obesity (Yes/No).

5. **Target Variable**
   - **Class**: Whether the patient is predicted to be Positive or Negative for diabetes.

Key Themes & Potential Questions

1. Demographic Insights

- What is the age and gender distribution of patients?
- Which age group or gender has a higher prevalence of diabetes?

2. Medical Indicator Analysis

- How do blood glucose, blood pressure, BMI, HbA1c differ in diabetic vs non-diabetic?
- Are there threshold levels that suggest high risk?

3. Symptom Pattern Analysis

- Which symptoms are most common in diabetic patients?
- What are the top co-occurring symptoms?

4. Comorbidity & Risk Factor Impact

- What % of diabetic patients also have hypertension, heart disease, or smoking history?
- How do multiple risk factors compound diabetes risk?

5. Class Prediction & Key Drivers

- What are the strongest predictors of a “Positive” diabetes class?
- Can a pattern of test results + symptoms predict diabetes?

6. Inter-feature Correlations

- How are BMI, BP, cholesterol, triglycerides, glucose interrelated?
- Do certain medical test pairs correlate with Class?

Symptom Clustering	80% of patients with Polyuria also have Polydipsia
Age Risk Profile	Diabetes most common in 50–65 age group
Gender Disparity	Male patients slightly more likely to test Positive
Medical Red Flags	Patients with BMI > 30 and Glucose > 150 are 3x more likely diabetic
Co-morbidity Risk	Smoking + Hypertension significantly increases diabetes risk

Thank You!