Development Part 1: Data Collection and Preprocessing

**1.Define the problem scope:**

\* Determine the specific type of diabetes you want to predict (e.g., Type 1 or Type 2).

\* Decide what features (attributes) you’ll use for prediction, such as age, BMI, family history, glucose levels, etc.

\* Understand the goal of the prediction (e.g., early diagnosis, risk assessment).

**2. Data collection:**

\* Gather a dataset containing historical health records of individuals, especially those diagnosed with diabetes.

\* Ensure your dataset is diverse, representative, and has a sufficient number of positive (diabetic) and negative (non-diabetic) cases.

3.Data preprocessing

\* Handle missing data by imputing or removing incomplete records.

\* Normalize or standardize features to ensure they have a consistent scale.

\* Encode categorical data (e.g., gender) into numerical format.

\* Split the dataset into training and testing sets for model evaluation.

Development Part 2: Model Building and Training

For this part, we will focus on data collection and preprocessing. You’ll typically use Python and relevant libraries like NumPy, Pandas, and Scikit-Learn.

Here’s some sample Python code to get you started with data preprocessing:

* **Python**

Import pandas as pd

From sklearn.model\_selection import train\_test\_split

From sklearn.preprocessing import StandardScaler

# Load your dataset

Data = pd.read\_csv(‘diabetes\_dataset.csv’) # Replace with your dataset file

# Handle missing data (e.g., filling missing values with the mean)

Data = data.fillna(data.mean())

# Split data into features (X) and target (y)

X = data.drop(‘diabetes\_status’, axis=1) # Replace ‘diabetes\_status’ with your target column

Y = data[‘diabetes\_status’]

# Normalize/standardize the features

Scaler = StandardScaler()

X = scaler.fit\_transform(X)

# Split the data into training and testing sets (e.g., 80% training, 20% testing)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)