# AI BASED DIABETES PREDICTION SYSTEM

## Phase 3: Development Part 1

**Introduction:**

The AI-based Diabetes Prediction System is a sophisticated application that utilizes artificial intelligence (AI) techniques to predict the likelihood of an individual developing diabetes. Diabetes is a chronic metabolic disorder characterized by high blood sugar levels, and early detection and intervention are crucial in managing the condition effectively. By employing advanced machine learning algorithms, this system can analyze various risk factors and provide accurate predictions, enabling healthcare professionals to take proactive measures and individuals to adopt preventive measures.

**How the AI-based Diabetes Prediction System Works:**

1. **Data Collection:** The system gathers a comprehensive set of data from individuals, which typically includes demographic information, medical history, lifestyle factors (such as diet and exercise), and relevant biomarkers (such as blood glucose levels, cholesterol levels, and body mass index).

2. **Data Preprocessing:** The collected data is then preprocessed to handle missing values, outliers, and standardize the variables. This step ensures that the data is in a suitable format for further analysis.

3. **Feature Selection:** The system identifies the most relevant features or variables that contribute significantly to diabetes prediction. This step helps in reducing the dimensionality of the data and improving the efficiency of the prediction model.

4. **Model Training:** The system employs machine learning algorithms, such as logistic regression, decision trees, random forests, or neural networks, to train a predictive model. During this process, the model

learns from the historical data to recognize complex patterns and relationships between the input variables and the occurrence of diabetes.

5. **Model Evaluation:** The trained model is evaluated using various performance metrics to assess its accuracy, such as sensitivity, specificity, precision, and area under the receiver operating characteristic curve (AUC-ROC).

6. **Prediction:** Once the model is trained and evaluated, it can make predictions on new, unseen data. By inputting relevant information about an individual, such as their demographics, medical history, and biomarkers, the system can provide a probability or risk score indicating the likelihood of developing diabetes within a certain timeframe.

**Benefits of the AI-based Diabetes Prediction System:**

1. **Early Detection:** By accurately predicting the risk of diabetes, the system enables early intervention, allowing healthcare professionals to initiate preventive measures and lifestyle modifications promptly.

2. **Personalized Recommendations:** Based on the predicted risk, the system can provide personalized recommendations to individuals, such as dietary modifications, exercise regimens, and regular health check-ups, to reduce the chances of developing diabetes.

3. **Healthcare Resource Optimization:** By identifying individuals at a high risk of diabetes, healthcare resources can be allocated more efficiently, focusing on targeted interventions and screenings for those who need them the most.

4. **Research and Insights:** The system's analysis of vast amounts of data can uncover valuable insights into the risk factors and causes of

diabetes, contributing to ongoing research efforts and improving our understanding of the disease.

**Program:**

```
import numpy as np
import pandas as pd
import tensorflow as tf
from keras.layers import Dense,Dropout
from sklearn.model_selection import train_test_split
import matplotlib as mlp
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn.preprocessing import StandardScaler
data=pd.read_csv("pima-indians-diabetes.csv")
data.head()
data = data.rename(index=str, columns={"6":"preg"})
data = data.rename(index=str, columns={"148":"gluco"})
data = data.rename(index=str, columns={"72":"bp"})
data = data.rename(index=str, columns={"35":"stinmm"})
data = data.rename(index=str, columns={"0":"insulin"})
data = data.rename(index=str, columns={"33.6":"mass"})
data =data.rename(index=str, columns={"0.627":"dpf"})
data = data.rename(index=str, columns={"50":"age"})
data = data.rename(index=str, columns={"1":"target"})
data.head()
X = data.iloc[:, :-1]
Y = data.iloc[:,8]
X_train_full, X_test, y_train_full, y_test = train_test_split(X, Y, random_state=42)
X_train, X_valid, y_train, y_valid = train_test_split(X_train_full, y_train_full,
random_state=42)
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
```

```
X_valid = scaler.transform(X_valid)
X_test = scaler.transform(X_test)
np.random.seed(42)
tf.random.set_seed(42)
 model=Sequential()
model.add(Dense(15,input_dim=8, activation='relu'))
model.add(Dense(10,activation='relu'))
model.add(Dense(8,activation='relu'))
model.add(Dropout(0.25))
model.add(Dense(1, activation='sigmoid'))
model.compile(loss="binary_crossentropy", optimizer="SGD",
metrics=['accuracy'])
model_history = model.fit(X_train, y_train, epochs=200,
validation_data=(X_valid, y_valid))
```

**Output:**

| | 6 | 148 | 72 | 35 | 0 | 33.6 | 0.627 | 50 | 1 |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 85 | 66 | 29 | 0 | 26.6 | 0.351 | 31 | 0 |
| **1** | 8 | 183 | 64 | 0 | 0 | 23.3 | 0.672 | 32 | 1 |
| **2** | 1 | 89 | 66 | 23 | 94 | 28.1 | 0.167 | 21 | 0 |
| **3** | 0 | 137 | 40 | 35 | 168 | 43.1 | 2.288 | 33 | 1 |
| **4** | 5 | 116 | 74 | 0 | 0 | 25.6 | 0.201 | 30 | 0 |

| | preg | gluco | bp | stinmm | insulin | mass | dpf | age | target |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 85 | 66 | 29 | 0 | 26.6 | 0.351 | 31 | 0 |
| 1 | 8 | 183 | 64 | 0 | 0 | 23.3 | 0.672 | 32 | 1 |
| 2 | 1 | 89 | 66 | 23 | 94 | 28.1 | 0.167 | 21 | 0 |
| 3 | 0 | 137 | 40 | 35 | 168 | 43.1 | 2.288 | 33 | 1 |
| 4 | 5 | 116 | 74 | 0 | 0 | 25.6 | 0.201 | 30 | 0 |

**Explanation:**

The code you provided is for training a diabetes prediction model using a neural network with TensorFlow and Keras. Here's a step-by-step explanation of the code:

1. **Importing the required libraries:**
    - `numpy` (as `np`): For numerical operations on arrays.
    - `pandas` (as `pd`): For data manipulation and analysis.
    - `tensorflow` (as `tf`): For building and training deep learning models.
    - `Dense` and `Dropout` from `keras.layers`: For defining the layers of the neural network.
    - `train_test_split` from `sklearn.model_selection`: For splitting the data into training and testing sets.
    - `matplotlib` (as `mlp`) and `matplotlib.pyplot` (as `plt`): For data visualization.

2. **Reading the data:** The code reads the data from a CSV file named "pima-indians-diabetes.csv" using `pd.read_csv` and assigns it to the `data` variable.

3. **Renaming the columns:** The code renames the columns of the `data` DataFrame to more descriptive names.

4. **Splitting the data into features and target variables:** The code separates the features (X) and the target variable (Y) from the `data` DataFrame.

5. **Splitting the data into training, validation, and testing sets:** The code uses `train_test_split` to split the data into training set (X_train, y_train), validation set (X_valid, y_valid), and testing set (X_test, y_test).

6. **Standardizing the features:** The code uses `StandardScaler` from `sklearn.preprocessing` to standardize the numerical features (X_train, X_valid, X_test).

7. **Defining the neural network model:** The code creates a sequential model using `Sequential` from `tensorflow.keras.models`. The model consists of several dense layers with ReLU activation and a dropout layer for regularization.

8. **Compiling the model:** The code compiles the model with a binary cross-entropy loss function, SGD optimizer, and accuracy metric.

**<u>Conclusion:</u>**

The AI-based Diabetes Prediction System is a powerful tool that leverages artificial intelligence and machine learning techniques to assess an individual's risk of developing diabetes. By providing accurate predictions and personalized recommendations, this system can play a significant role in preventive healthcare, early detection, and effective management of diabetes, ultimately improving health outcomes for individuals at risk.