**PHASE 04**

**AI-BASED DIABETES PREDICTION SYSTEM**

**INTRODUCTION:**

In this we will continue building our project by selecting a machine learning algorithm, training the model, and evaluating its performance. Let us see in detail.

**KEY CONCEPTS:**

Here we are using Logistic Regression for diabetes prediction. Let’s briefly discuss about key concepts related to logistic regression and diabetes prediction.

**Logistic Regression:**

Logistic regression is a statistical model used to predict the probability of a binary outcome. It is commonly used in machine learning for classification tasks. In our case, we will use logistic regression to predict whether a person has diabetes or not based on certain input features.

**Diabetes Prediction:**

Diabetes is a chronic condition that affects the body’s ability to regulate blood sugar levels. Predicting diabetes can be crucial for early detection and effective management of the disease. Machine learning algorithms, such as logistic regression, can be used to analyze relevant features and make accurate predictions.

**CODE STRUCTURE:**

Data Preparation: Splitting the dataset into input features (X) and the target variable (y).

Train-Test Split: Splitting the data into training and testing sets.

Model Training: Creating a logistic regression model and fitting it to the training data.

Prediction: Using the trained model to predict the outcome for the test data.

Evaluation: Calculating the accuracy of the model and creating a confusion matrix.

Additional Prediction: Making a prediction for a new set of input features.

**CODE:**

#Logistic Regression

y = dataset\_new['Outcome']  
X = dataset\_new.drop('Outcome', axis=1)

#splitting X and Y  
from sklearn.model\_selection import train\_test\_split  
X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X, y, test\_size = 0.20, random\_state = 42, stratify = dataset\_new['Outcome'] )  
from sklearn.linear\_model import LogisticRegression  
model = LogisticRegression()  
model.fit(X\_train, Y\_train)  
y\_predict = model.predict(X\_test)  
y\_predict

#confusion matrix  
from sklearn.metrics import confusion\_matrix  
cm = confusion\_matrix(Y\_test, y\_predict)  
cm

#heatmap of confusion matrix  
sns.heatmap(pd.DataFrame(cm), annot=True)  
from sklearn.metrics import accuracy\_score  
accuracy =accuracy\_score(Y\_test, y\_predict)  
accuracy  
y\_predict = model.predict([[1,148,72,35,79.799,33.6,0.627,50]])  
print(y\_predict)  
if y\_predict==1:  
    print("Diabetic")  
else:  
    print("Non Diabetic")

**CONCLUSION:**

In this we learned how to use logistic regression to predict diabetes. Logistic regression is a powerful algorithm for binary classification tasks, and it can be applied to various domains, including healthcare. By understanding the concept we can build our own diabetes prediction models using logistic regression.