PHASE 03

INTRODUCTION:

We are assigned with the task of designing an AI-Based Diabetes Prediction System.

DATASET:

In this document we guys are here to discuss the loading and preprocessing of the dataset to make a better prediction for diabetes.

PYTHON PROGRAM:

**#Importing libraries**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

**#Importing dataset**

dataset = pd.read\_csv(‘diabetes.csv’)

dataset.head()

dataset.shape

dataset.info()

dataset.describe().T

dataset.isnull().sum()

**#Data Visualization**

sns.countplot(x = 'Outcome',data = dataset)

sns.pairplot(data = dataset, hue = ‘Outcome’)

plt.show()

sns.heatmap(dataset.corr(), annot = True)

plt.show()

**#Processing the Data**

Dataset\_new = dataset

Dataset\_new[[“Glucose”, “BloodPressure”, “SkinThickness”, “Insulin”, “BMI”]] = dataset\_new[[“Glucose”, “BloodPressure”, “SkinThickness”, “Insulin”, “BMI”]].replace(0, np.NaN)

Dataset\_new.isnull().sum()

Dataset\_new[“Glucose”].fillna(dataset\_new[“Glucose”].mean(), inplace = True)

Dataset\_new[“BloodPressure”].fillna(dataset\_new[“BloodPressure”].mean(), inplace = True)

Dataset\_new[“SkinThickness”].fillna(dataset\_new[“SkinThickness”].mean(), inplace = True)

Dataset\_new[“Insulin”].fillna(dataset\_new[“Insulin”].mean(), inplace = True)

Dataset\_new[“BMI”].fillna(dataset\_new[“BMI”].mean(), inplace = True)

Dataset\_new.isnull().sum()