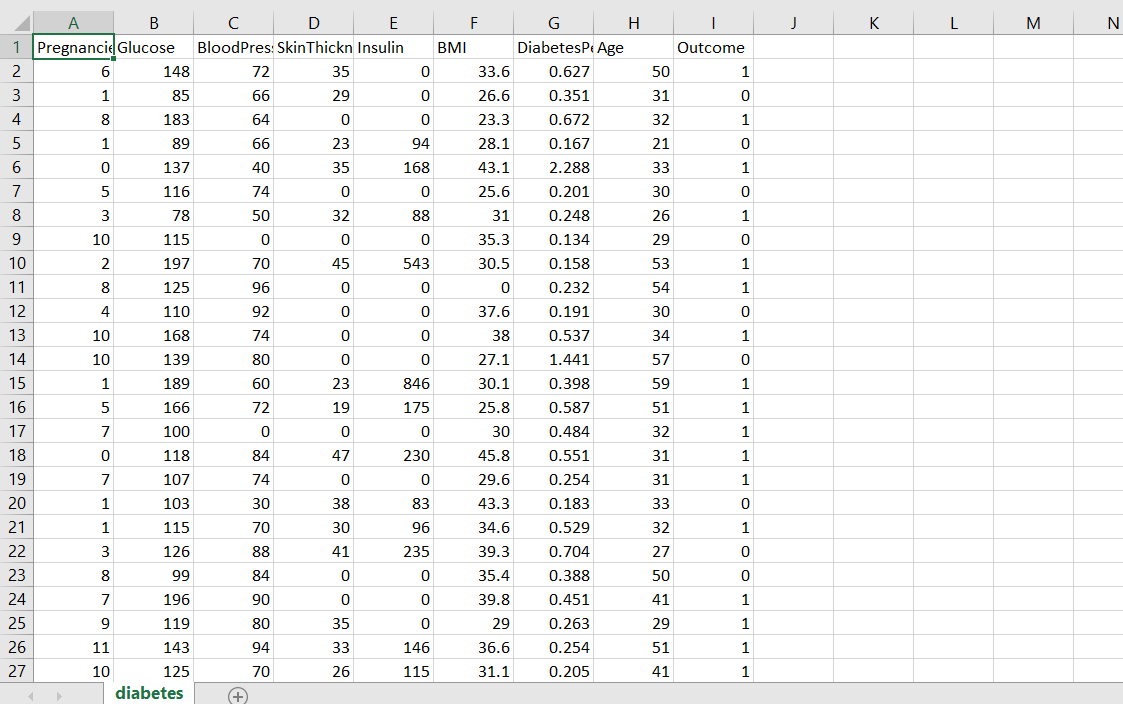
Ai-based diabetes prediction system

AI\_phase3

Development part 1

Document submission

DATASET OF DIABETES



INTRODUCTION:

Artificial intelligence (AI) is rapidly advancing and finding its application in various fields, including healthcare. One of the significant areas where AI is being utilized is in diabetes prediction and management. An AI-based diabetes prediction system utilizes advanced machine learning algorithms to analyze various factors and predict the likelihood of an individual developing diabetes.

This predictive system collects data from individuals' medical history, such as age, gender, weight, blood pressure, family history of diabetes, lifestyle choices, and other relevant health information. It then uses this data to build a model that can predict the probability of an individual developing diabetes in the future.

The AI algorithms used in diabetes prediction systems are trained on large datasets containing information about individuals who have been diagnosed with diabetes or are at risk. This training enables the model to recognize patterns and correlations between the given factors and the development of diabetes.

Once the AI model is trained, it can make accurate predictions about an individual's future risk of developing diabetes. The system can provide personalized recommendations and preventive measures that can help the individual reduce their risk or manage their condition more effectively. For instance, it can suggest lifestyle modifications, dietary changes, regular exercise routines, or even medication if necessary.

By providing early warnings and proactive measures, AI-based diabetes prediction systems can facilitate timely intervention, allowing individuals to adopt healthier habits and potentially delay or prevent the onset of diabetes. This technology has the potential to revolutionize diabetes management by empowering both patients and healthcare providers with valuable insights and personalized care plans.

However, it is important to note that AI-based diabetes prediction systems are still evolving and require continuous improvement and validation. The accuracy of the predictions depends on the quality and quantity of the data used to train the models. Additionally, these systems can never replace professional medical advice and should be seen as a tool to assist healthcare professionals in delivering personalized care.

To build an AI-based diabetes prediction system,

1.Import the necessary libraries:

python

import pandas as pd

import numpy as np

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

from sklearn.preprocessing import StandardScaler

2. Load the dataset:

python

df = pd.read\_csv('diabetes\_dataset.csv') # Replace 'diabetes\_dataset.csv' with your dataset file name

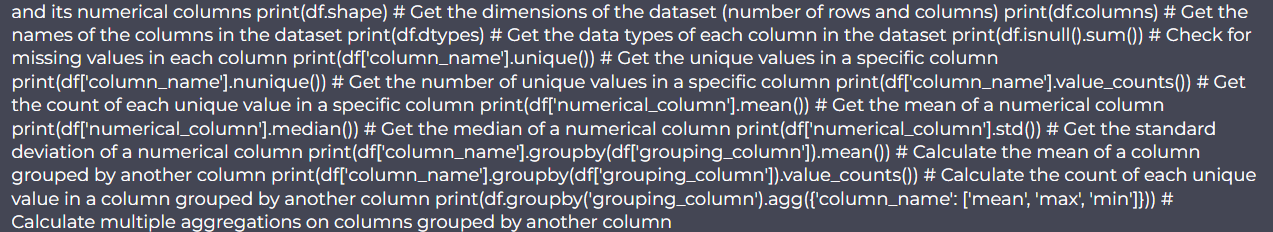
3. Explore the dataset:

python

print(df.head()) # Print the first few rows of the dataset to check the data

print(df.info()) # Get information about the dataset, such as column names and data types

print(df.describe()) # Get statistical summary of the dataset



4. Preprocess the dataset:

python

# Split the dataset into features (X) and target variable (y)

X = df.drop('diabetes\_class', axis=1)

y = df['diabetes\_class']

# Split the data into training and testing sets

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

# Apply feature scaling to normalize the feature values

scaler = StandardScaler()

X\_train = scaler.fit\_transform(X\_train)

X\_test = scaler.transform(X\_test)

5. Perform other analysis as needed, such as data visualization, correlation analysis , or feature selection. Example code snippets are provided below:

- Data visualization using matplotlib:

python

import matplotlib.pyplot as plt

# Boxplot to visualize distribution of all features

df.boxplot()

plt.show()

# Histogram to visualize the distribution of a feature (e.g., glucose)

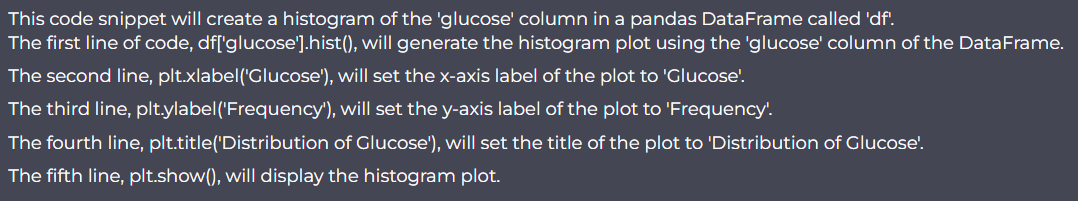
df['glucose'].hist()

plt.xlabel('Glucose')

plt.ylabel('Frequency')

plt.title('Distribution of Glucose')

plt.show()



- Correlation matrix heatmap using seaborn:

python

import seaborn as sns

# Compute the correlation matrix

corr\_matrix = df.corr()

# Create a heatmap of the correlation matrix

sns.heatmap(corr\_matrix, annot=True, cmap='coolwarm')

plt.title('Correlation Matrix Heatmap')

plt.show()

- Feature selection using scikit-learn (e.g., SelectKBest):

python

from sklearn.feature\_selection import SelectKBest, f\_classif

# Create the selector object to select the top k features

k = 5 # Select the top 5 features

selector = SelectKBest(score\_func=f\_classif, k=k)

# Fit the selector to the training data

X\_train\_selected = selector.fit\_transform(X\_train, y\_train)

# Get the indices of the selected features

selected\_indices = selector.get\_support(indices=True)

# Get the names of the selected features

selected\_features = X.columns[selected\_indices]

print(selected\_features)

These are just a few examples of the analysis you can perform on the dataset. Feel free to explore and experiment with other analysis techniques based on your specific requirements for the diabetes prediction system.