AI BASED DIABETES PREDICTION SYSTEM

TEAM MEMBER

**815821104035:YOGESH RAJ.R**

**PHASE 2 DOCUMENT SUBMISSION**

**PROJECT:DIABETES PREDICTION SYSTEM**

****

INTRODUCTION:

Diabetes is a global health epidemic that affects millions of people worldwide, and its prevalence continues to rise. According to the International Diabetes Federation, approximately 463 million adults had diabetes in 2019, and this number is expected to increase to 700 million by 2045. Managing diabetes and its complications is a significant challenge for both individuals and healthcare systems.

Artificial Intelligence (AI) has emerged as a powerful tool in various fields, including healthcare. AI-based systems are increasingly being employed to improve disease prevention, diagnosis, and management. In the context of diabetes, AI has the potential to revolutionize the way we predict and manage this chronic condition.

This introduction will provide an overview of an AI-based diabetes prediction system, explaining its significance, purpose, and the technology involved.

**1.Significance of Diabetes Prediction:**

Diabetes is a complex and multifactorial disease, making early detection and proactive management crucial to preventing complications such as heart disease, kidney failure, and vision problems. AI-based diabetes prediction systems offer a proactive approach by identifying individuals at risk of developing diabetes or those already living with the condition. This early identification allows for timely interventions, lifestyle modifications, and personalized treatment plans.

**2.Purpose of the AI-Based Diabetes Prediction System:**

a. Early Detection

b. Risk Stratificationc

c. Data-Driven Insights

d. Treatment Optimization

**3.Technology Involved**:

AI-based diabetes prediction systems employ a combination of technologies, including:

a. Machine Learning Algorithms.

b. Big Data Analysis

c. Natural Language Processing (NLP)

d. Continuous Monitoring.

**DATA SOURCE:**

Dataset Link: <https://www.kaggle.com/datasets/mathchi/diabetes-data-set>

An AI-based diabetes prediction system can be a valuable innovation in healthcare. Such a system could use machine learning and predictive analytics to identify individuals at risk of developing diabetes, allowing for early intervention and personalized healthcare. Here's a high-level overview of how you can create an innovative AI-based diabetes prediction system:

Data Collection and Integration:

Gather relevant healthcare data, which may include medical records, lifestyle information, genetic data, and wearable device data (such as fitness trackers and continuous glucose monitors). Integrating data from various sources is critical for comprehensive prediction.

Data Preprocessing:

Clean and preprocess the data to remove noise, handle missing values, and standardize formats. This step is crucial to ensure data quality.

Feature Selection and Engineering:

Identify relevant features (variables) for the prediction model. Feature engineering may involve creating new variables that better represent underlying patterns related to diabetes risk.

Machine Learning Model Selection:

Choose appropriate machine learning algorithms for prediction, such as logistic regression, decision trees, random forests, or more advanced techniques like deep learning (e.g., neural networks).

Model Training:

Train the selected model using historical data, where the outcome variable is the presence or absence of diabetes. You can use techniques like cross-validation to assess model performance.

PROGRAM:

Importing Libraries

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

sns.set()

from mlxtend.plotting import plot\_decision\_regions

import missingno as msno

from pandas.plotting import scatter\_matrix

from sklearn.preprocessing import StandardScaler

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

from sklearn.neighbors import KNeighborsClassifier

from sklearn.metrics import confusion\_matrix

from sklearn import metrics

from sklearn.metrics import classification\_report

import warnings

warnings.filterwarnings('ignore')

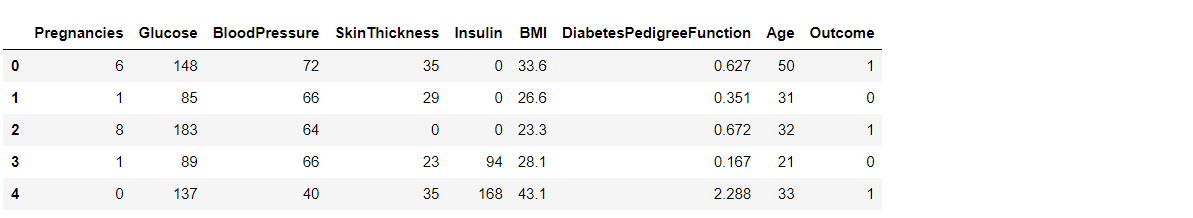
%matplotlib inline

Here we will be reading the dataset which is in the CSV format

diabetes\_df = pd.read\_csv('diabetes.csv')

diabetes\_df.head()

**Output:**



**Exploratory Data Analysis (EDA**):

Now let’ see that what are columns available in our dataset.

diabetes\_df.columns

**Output:**

Index(['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin',

'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome'],

dtype='object')

Information about the dataset

diabetes\_df.info()

**Output:**

RangeIndex: 768 entries, 0 to 767

Data columns (total 9 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Pregnancies 768 non-null int64

1 Glucose 768 non-null int64

2 BloodPressure 768 non-null int64

3 SkinThickness 768 non-null int64

4 Insulin 768 non-null int64

5 BMI 768 non-null float64

6 DiabetesPedigreeFunction 768 non-null float64

7 Age 768 non-null int64

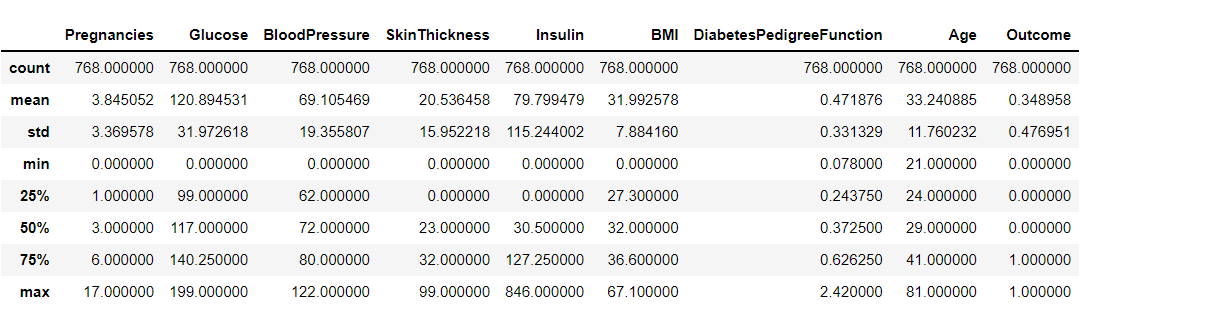
8 Outcome 768 non-null int64

dtypes: float64(2), int64(7)

To know more about the dataset

diabetes\_df.describe()

**output:**

****

**CONCLUSION:**

In diabetes diagnosis and treatment, AI-based medical devices have already been approved by the FDA and are available in other countries as well. Currently, many studies have used machine learning to predict the onset of diabetes. However, these machine learning approaches have not demonstrated superior performance in predicting disease onset compared to conventional statistical techniques that combine risk factors**.**