AI BASED DIABETES PREDICTION SYSTEM

TEAM MEMBER:

NAME: M.RUDHRESHKUMAR

REGISTER NO: 410121243045

DOCUMENT SUBMISSION

PHASE 4 ---

PROJECT OVERVIEW

Introduction:

Diabetes is a chronic medical condition that affects millions of people worldwide. Early detection and management of diabetes are crucial for improving the quality of life and reducing complications. This project aims to develop an AI-based Diabetes Prediction System that can predict the risk of diabetes in individuals based on their health data and provide early warnings and recommendations.

Problem Statement:

The primary objective of this project is to create a predictive model that can analyze various health-related data points and provide an accurate prediction of an individual's risk of developing diabetes. The system will consider a range of risk factors, including age, gender, family history, lifestyle, and medical history, to provide personalized predictions.

Key Features and Components:

DATA COLLECTION:

Gather a diverse dataset of health records, including patient demographics, medical history, lifestyle data, and biomarker information.

DATA PROCESSING:

Clean and preprocess the data, handling missing values and outliers, and normalizing/standardizing features.

FEATURE SELECTION:

Identify and select the most relevant features that contribute to diabetes risk prediction.

MACHINE LEARNING MODELS :

Develop and train machine learning models, such as logistic regression, decision trees, random forests, or deep learning models, using the preprocessed data.

EVALUATION METRICS:

Implement evaluation metrics (e.g., accuracy, F1 score, ROC-AUC) to assess the performance of the predictive model.

USER INTERFACE:

Create a user-friendly interface for users to input their health data and receive risk predictions.

PERSONALISED RECOMMENDATIONS:

Provide personalized recommendations based on risk levels, such as lifestyle changes, diet modifications, and exercise routines.

PRIVACY AND SECURITY:

Implement strong security measures to protect user health data and ensure compliance with data privacy regulations.

BENEFITS:

Early Detection:

The system will assist in identifying individuals at risk of diabetes at an early stage, enabling timely intervention and treatment.

Personalized Recommendations:

Users will receive tailored recommendations to reduce their diabetes risk, promoting a healthier lifestyle.

Health Monitoring:

The system can be used for continuous health monitoring and provide regular updates on diabetes risk.

Target Audience:

Healthcare professionals for clinical use.

Individuals interested in monitoring their diabetes risk.

Future Enhancements:

Integration with wearable devices for real-time health data collection.

Expansion to include additional chronic diseases for prediction.

Collaboration with healthcare institutions for data sharing and research.

The AI-Based Diabetes Prediction System aims to leverage the power of artificial intelligence and machine learning to provide accurate and personalized predictions for diabetes risk. By implementing this system, individuals can take proactive steps towards preventing or managing diabetes, ultimately improving their overall health and well-being.

Program code:

```
# Import necessary libraries

import numpy as np
```

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report

# Load the diabetes dataset from scikit-learn
from sklearn.datasets import load_diabetes
data = load_diabetes()
X = data.data
y = (data.target > 140).astype(int)  # Binary classification: 1 if diabetes, 0 if not

# 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)

# Standardize the feature data
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Create and train a Logistic Regression model
model = LogisticRegression()
model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = model.predict(X_test)

# Calculate and print the accuracy
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy:.2f}")
```

```
# Print a classification report
report = classification_report(y_test, y_pred)
print("Classification Report:\n", report)
```

OUTPUT:

Accuracy: 0.72

Classification Report:

	precision	recall	f1-score	support
0	0.78	0.72	0.75	39
1	0.64	0.71	0.67	29
accuracy			0.72	68
macro avg	0.71	0.72	0.71	68
weighted avg	0.72	0.72	0.72	68

annexure

Program (code):

```
# Import necessary libraries
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix

# Load the diabetes dataset
```

```
from sklearn.datasets import load_diabetes

diabetes = load_diabetes()


# Create a DataFrame from the dataset

data = pd.DataFrame(data=diabetes.data, columns=diabetes.feature_names)

target = pd.Series(diabetes.target, name='target')

data['target'] = target


# Split the data into features (X) and target (y)

X = data.drop('target', axis=1)

y = data['target']


# 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)


# Create a logistic regression model

model = LogisticRegression()


# Fit the model on the training data

model.fit(X_train, y_train)


# Make predictions on the test data

y_pred = model.predict(X_test)


# Calculate accuracy

accuracy = accuracy_score(y_test, y_pred)


# Print the accuracy and confusion matrix

print("Accuracy:", accuracy)

print("Confusion Matrix:")

print(confusion_matrix(y_test, y_pred))
```

INPUT:

She patient is likely does not have diabetes.

OUTPUT:

Pregnancies: 2

Glucose: 120

BloodPressure: 70

SkinThickness: 30

Insulin: 80

BMI: 25

DiabetesPedigreeFunction: 0.4

Age: 35