AI BASED DIABETES PREDICTION SYSTEM

Problem Definition:

A dataset of individuals, including features such as age, sex, BMI, blood pressure, blood sugar levels, family history of diabetes, lifestyle factors (e.g., diet, exercise), and other relevant medical information.An accurate and reliable diabetes prediction system could help individuals identify their risk of developing the disease early on.

Design Thinking:

1.Empathize: Understand the needs of the users. Who are the users of the diabetes prediction system? What are their goals? What are their pain points. \*Understand the challenges of diabetes. What are the different factors that can contribute to diabetes? What are the risks of developing diabetes?

2.Define:

Identify the key problems that the system should address.What are the most important factors to consider when predicting diabetes risk? What level of accuracy is needed? Define the desired outcomes of the system. What should users be able to do with the system? How should it help them to manage their diabetes risk?

3.Ideate:

Brainstorm different ways to address the key problems. What different data sources could be used? What different machine learning algorithms could be employed? Consider the different needs of different users. How can the system be designed to be accessible and useful to everyone?

4.Prototype:

Build a prototype of the system to test out different ideas. This could be a simple web app or a more sophisticated mobile app. Get feedback from users to refine the prototype. What do users like? What do they dislike? What features would they like to see added?

5.Test:

Test the prototype with a larger group of users to evaluate its performance.How accurate is the system’s predictions? Is it easy to use? Is it useful to users? Make necessary adjustments to the system based on the feedback.

6.Deploy:

Deploy the system to a wider audience.Make it available to users through a web app, mobile app, or other platform.Continue to monitor the system’s performance and collect feedback from users. This will help to ensure that the system is meeting the needs of users and that it is providing accurate and reliable predictions.

AI-Based Diabetes Prediction System

An AI-based diabetes prediction system can be used to identify individuals at high risk of developing diabetes, enabling early intervention and prevention. The system can be trained on a large dataset of electronic health records, including patient demographics, medical history, and laboratory results. The trained model can then be used to predict the risk of developing diabetes for new patients.

The system can be used In a variety of settings, including primary care clinics, hospitals, and public health departments. It can also be integrated into wearable devices and smartphone apps, allowing individuals to track their own risk factors and receive personalized recommendations to reduce their risk of diabetes.

Steps to Transform the Design into Reality

To transform the design of an AI-based diabetes prediction system into reality, the following steps can be taken:

1. Collect data: A large dataset of electronic health records is needed to train the AI model. This data can be collected from primary care clinics, hospitals, and other healthcare providers.

2. Clean and prepare the data: The data must be cleaned and prepared before it can be used to train the model. This involves removing errors and inconsistencies in the data, and converting the data into a format that the model can understand.

3. Choose an AI algorithm: There are a variety of AI algorithms that can be used for diabetes prediction. Some popular algorithms include logistic regression, support vector machines, and random forests.

4. Train the model: The AI algorithm is trained on the prepared data. During training, the model learns to identify patterns in the data that are associated with diabetes risk.

5.Evaluate the model: Once the model is trained, it must be evaluated on a held-out test set. This helps to ensure that the model is generalizable to new data and does not overfit the training data.

6. Deploy the model: Once the model is evaluated and found to be accurate, it can be deployed to production. This may involve integrating the model into a clinical information system, wearable device, or smartphone app.

Challenges and Considerations

There are a number of challenges and considerations that must be taken into account when developing and deploying an AI-based diabetes prediction system. These include:

1.Data privacy and security: The data used to train and deploy the model must be protected from unauthorized access and use.

2.Model interpretability: It is important to be able to explain how the model makes predictions, so that clinicians and patients can understand and trust the system.

3.Model bias: The model must be trained on a representative dataset that reflects the diversity of the population it will be used to predict diabetes risk for.

4.Clinical integration: The system must be integrated into existing clinical workflows in order to be used by clinicians and patients.

Despite these challenges, AI-based diabetes prediction systems have the potential to revolutionize the way that diabetes is prevented and managed. By identifying individuals at high risk of developing diabetes, these systems can enable early intervention and improve patient outcomes.

Program Code:

Import numpy as np

Import pandas as pd

Import matplotlib.pyplot as plt

Import seaborn as sns

Dataset=pd.read\_csv(‘C:\\Users\\STUDENT\\Desktop\\diabetes.csv’)

Dataset.head()

Dataset.shape

Dataset.isnull().values.any()

Dataset.info()

Dataset.describe()

Dataset.isnull().sum()

Sns.countplot(x = ‘Outcome’,data = dataset)

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

Plt.show()

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

Plt.show()

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()

# Check for Missing Values

Missing\_values = df.isnull().sum()

Print(“Missing Values:”)

Print(missing\_values)

# Handle missing values (if any)

# For example, fill missing values with the mean of the column

Mean\_fill = df.mean()

Df.fillna(mean\_fill, inplace=True)

# Check for Duplicate Rows

Duplicate\_rows = df[df.duplicated()]

Print(“\nDuplicate Rows:”)

Print(duplicate\_rows)

# Handle duplicate rows (if any)

# For example, drop duplicate rows

Df.drop\_duplicates(inplace=True)

# Step 4: Data Analysis

# Summary Statistics

Summary\_stats = df.describe()

Print(“\nSummary Statistics:”)

Print(summary\_stats)

# Class Distribution (for binary classification problems)

Class\_distribution = df[‘Outcome’].value\_counts()

Print(“\nClass Distribution:”)

Print(class\_distribution)

# Step 5: Data Visualization

Sns.pairplot(df, hue=’Outcome’)

Plt.show()

# Step 6: Support Vector Machine (SVM) Modeling

# Separate features and target variable

X = df.drop(‘Outcome’, axis=1)

Y = df[‘Outcome’]

# Split the dataset into a training and testing set

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

# Standardize features

Scaler = StandardScaler()

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

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

# Initialize and train the SVM model

Model = SVC(kernel=’linear’, random\_state=42)

Model.fit(X\_train, y\_train)

# Make predictions

Y\_pred = model.predict(X\_test)

# Evaluate the model

Accuracy = accuracy\_score(y\_test, y\_pred)

Print(f’Accuracy: {accuracy:.2f}’)

# Classification report and confusion matrix

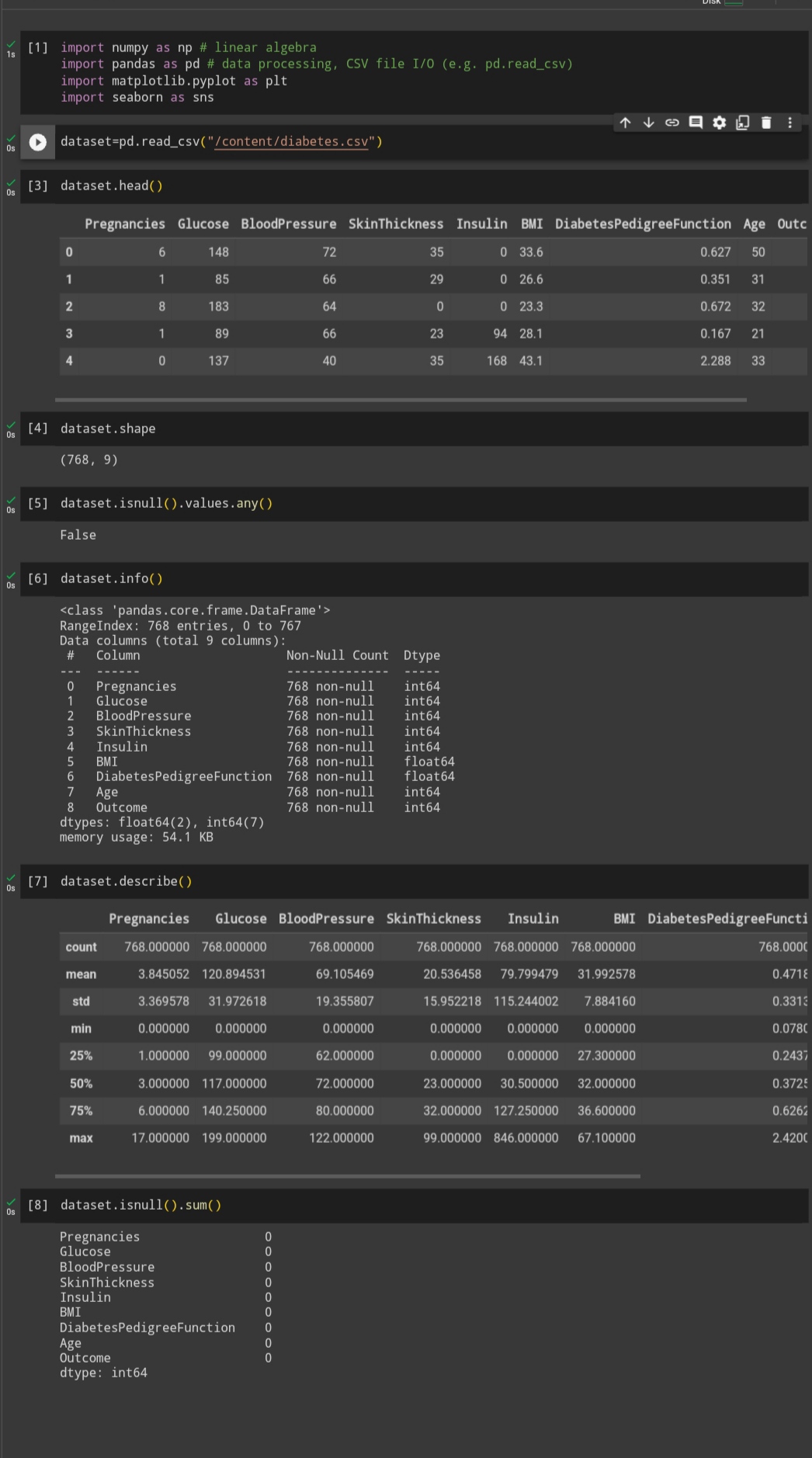
Print(classification\_report(y\_test, y\_pred))

Cm = confusion\_matrix(y\_test, y\_pred)

Sns.heatmap(cm, annot=True, fmt=’d’)

Plt.show()

OUTPUT:



Missing Values:

Pregnancies 0

Glucose 0

BloodPressure 0

SkinThickness 0

Insulin 0

BMI 0

DiabetesPedigreeFunction 0

Age 0

Outcome 0

Dtype: int64

Duplicate Rows:

Empty DataFrame

Columns: [Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, Outcome]

Index: []

Summary Statistics:

Pregnancies Glucose BloodPressure SkinThickness Insulin \

Count 768.000000 768.000000 768.000000 768.000000 768.000000

Mean 3.845052 120.894531 69.105469 20.536458 79.799479

Std 3.369578 31.972618 19.355807 15.952218 115.244002

Min 0.000000 0.000000 0.000000 0.000000 0.000000

25% 1.000000 99.000000 62.000000 0.000000 0.000000

50% 3.000000 117.000000 72.000000 23.000000 30.500000

75% 6.000000 140.250000 80.000000 32.000000 127.250000

Max 17.000000 199.000000 122.000000 99.000000 846.000000

BMI DiabetesPedigreeFunction Age Outcome

Count 768.000000 768.000000 768.000000 768.000000

Mean 31.992578 0.471876 33.240885 0.348958

Std 7.884160 0.331329 11.760232 0.476951

Min 0.000000 0.078000 21.000000 0.000000

25% 27.300000 0.243750 24.000000 0.000000

50% 32.000000 0.372500 29.000000 0.000000

75% 36.600000 0.626250 41.000000 1.000000

Max 67.100000 2.420000 81.000000 1.000000

Class Distribution:

Outcome

1. 500
2. 1 268

Name: count, dtype: int64



Accuracy: 0.76

Precision recall f1-score support

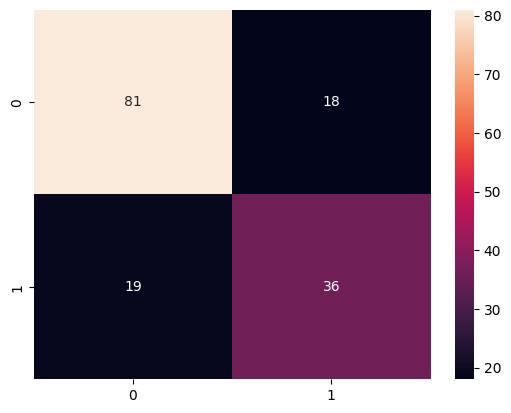
0 0.81 0.82 0.81 99

1 0.67 0.65 0.66 55

Accuracy 0.76 154

Macro avg 0.74 0.74 0.74 154

Weighted avg 0.76 0.76 0.76 154



Dataset: Diabetes prediction Dataset is provided

Features Used in dataset:

AI diabetes prediction system databases typically contain a variety of features, including:

\* \*\*Demographic features:\*\* These features include the patient’s age, sex, race, and ethnicity.

\* \*\*Medical history features:\*\* These features include the patient’s history of diabetes, prediabetes, and other medical conditions, such as heart disease, stroke, and high blood pressure.

\* \*\*Lifestyle features:\*\* These features include the patient’s weight, height, body mass index (BMI), diet, exercise habits, and smoking status.

\* \*\*Laboratory data:\*\* This data may include blood glucose levels, cholesterol levels, blood pressure, and other laboratory results.

In addition to these core features, some AI diabetes prediction system databases may also contain other features, such as:

\* \*\*Genetic data:\*\* This data can be used to identify genetic risk factors for diabetes.

\* \*\*Wearable device data:\*\* This data can be collected from wearable devices, such as smartwatches and fitness trackers, and can include information such as heart rate, sleep patterns, and activity levels.

\* \*\*Social media data:\*\* This data can be collected from social media platforms, such as Twitter and Facebook, and can be used to identify factors such as social isolation and stress levels, which may be associated with an increased risk of diabetes.

By combining these different types of data, AI diabetes prediction system databases can be used to develop models that can accurately predict a patient’s risk of developing diabetes.

Here are some specific examples of features that might be found in an AI diabetes prediction system database:

\* Age: Age is a strong risk factor for diabetes. The risk of developing diabetes increases with age, especially after the age of 45.

\*Sex: Women are more likely to develop diabetes than men.

\*Race and ethnicity: African Americans, Hispanics/Latinos, American Indians, and Asian Americans are more likely to develop diabetes than Caucasians.

\*Family history: People with a family history of diabetes are more likely to develop the disease themselves.

\*Body mass index (BMI):People who are overweight or obese are more likely to develop diabetes.

\*Blood pressure:High blood pressure is a risk factor for diabetes.

\*Cholesterol levels: High cholesterol levels are a risk factor for diabetes.

\*Blood sugar levels:High blood sugar levels are a sign of diabetes.

Other features that might be found in an AI diabetes prediction system database include:

\*Medications:The types and doses of medications that the patient is taking.

\*Diet:The patient’s typical diet, including the types and quantities of food and beverages consumed.

\*Exercise:The patient’s exercise habits, including the frequency, intensity, and duration of exercise.

\*Smoking status: Whether or not the patient smokes.

\*Social media data: The patient’s social media activity, such as the number of friends they have, the types of posts they share, and the people they interact with.

By combining these different types of data, AI diabetes prediction systems can be used to develop models that can accurately predict a patient’s risk of developing diabetes. This information can then be used to help patients take steps to prevent or manage the disease.

Conclusion

\*AI-based diabetes prediction systems have the potential to play a significant role in the prevention and management of diabetes. By identifying individuals at high risk of developing diabetes, these systems can enable early intervention and improve patient outcomes.

\*To transform the design of an AI-based diabetes prediction system into reality, a number of steps must be taken, including collecting and preparing data, choosing an AI algorithm, training and evaluating the model, and deploying the model.

\*It is important to address the challenges and considerations associated with AI-based diabetes prediction systems, such as data privacy and security, model interpretability, model bias, and clinical integration.

\*Overall, AI-based diabetes prediction systems have the potential to make a significant positive impact on public health.