**PHASE-3 : AI Based Diabetes Prediction System:**

**MODEL FOR innovation for an AI Diabetes Prediction System**:

Imagine a healthcare technology company, HealthTech Innovators, embarking on the development of an AI-based Diabetes Prediction System. Here’s how they apply the model for innovation:

**Obtaining the Dataset:**

* The first step in any machine learning project is to obtain a dataset to work with. There are many online resources available to find the datasets suitable for our project. One such popular platform that hosts datasets is Kaggle. Since the goal of our project is to predict Diabetes, I search for diabetes patient’s dataset, which is readily available as were already many such similar projects.
* Go to the Kaggle dataset page you mentioned: https://www.kaggle.com/datasets/mathchi/diabetes-data-set
* Click the “Download” button to get the dataset files.
* Unzip the downloaded files to a directory on your local machine into your project repository
* This dataset has the following attributes : Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function and Age, It also contains the Outcome label which tells whether the patient has diabetes or not. Since there is a class label present by default, this dataset can be studied using Classification based algorithms.

**Problem Identification:**

* Recognize the global diabetes epidemic as a pressing issue affecting millions.
* Identify the need for an AI system that can predict diabetes risk and empower healthcare professionals and individuals.

**Multidisciplinary Team:**

* Form a team consisting of data scientists, endocrinologists, nutritionists, user experience designers, and cybersecurity experts.

**Continuous Research**:

* Conduct extensive research to understand the complexities of diabetes, its risk factors, and potential complications.

**Ethical Data Handling:**

* Develop strict data privacy protocols to ensure compliance with healthcare data regulations.
* Implement robust encryption and access control mechanisms.

**Machine Learning Expertise:**

* Utilize cutting-edge machine learning algorithms and collaborate with AI researchers to enhance prediction accuracy.

**Design Thinking Approach:**

* Empathize with healthcare providers and individuals through interviews and surveys to identify their specific needs and concerns.

**Prototyping and Usability Testing:**

* Create interactive prototypes of the AI system’s user interface.
* Conduct usability testing with healthcare professionals and individuals to refine the design.

**Real-World Deployment:**

* Roll out the AI system in partnership with healthcare institutions and clinics.
* Provide comprehensive training to healthcare staff and user-friendly interfaces for individuals.

**Continuous Monitoring:**

* Set up a monitoring system to track prediction accuracy and system performance.
* Regularly gather user feedback and iterate on system improvements.

**Scalability and Accessibility:**

* Design the system to handle a growing volume of health data.
* Ensure the system is accessible to individuals of diverse backgrounds and regions.

**Collaboration with Healthcare Institutions:**

* Partner with leading healthcare institutions to integrate the AI system into their electronic health record systems.
* Collaborate on research initiatives to advance diabetes care.

**CODE:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

dataset = pd.read\_csv('Documents/Roshan/diabetes.csv') dataset.head()

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Pregnancies | Glucose | BloodPressure | SkinThickness |  | Insulin | BMI | DiabetesPedigreeFunction | Age | Outcome |
| 0 6 | 148 | 72 | 35 |  | 0 | 33.6 | 0.627 | 50 | 1 |
| 1 1 | 85 | 66 | 29 |  | 0 | 26.6 | 0.351 | 31 | 0 |
| 2 8 | 183 | 64 | 0 |  | 0 | 23.3 | 0.672 | 32 | 1 |
| 3 1 | 89 | 66 | 23 |  | 94 | 28.1 | 0.167 | 21 | 0 |
| 4 0 | 137 | 40 | 35 |  | 168 | 43.1 | 2.288 | 33 | 1 |

dataset.shape

(768, 9)

dataset.info()

‹class 'pandas.core.frame.DataFrame'> RangeIndex: 768 entries, 0 to 767 Data columns (total 9 columns):

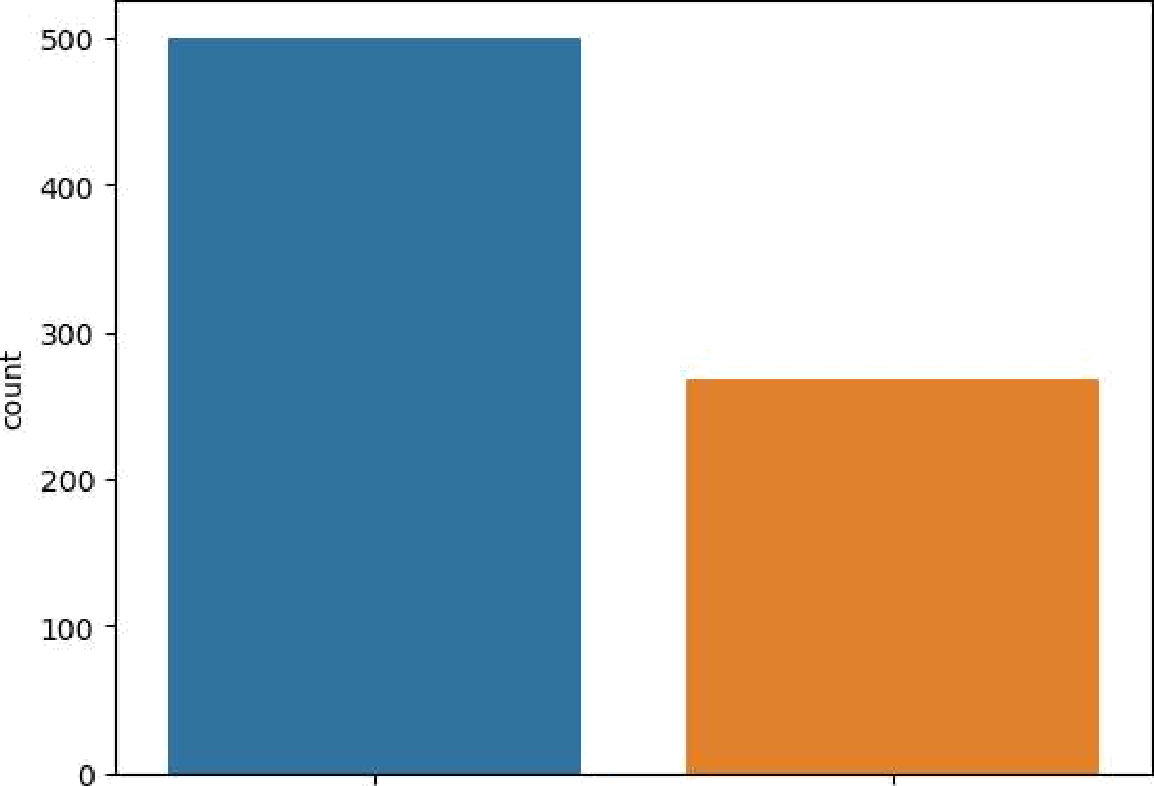
Non-Null Count Dtype

|  |  |  |  |
| --- | --- | --- | --- |
| 0 | Pregnancies | 768 non-null | iut64 |
| 1 | Glucose | 768 non-null | int64 |
| 2 | BloodPressure | 768 non-null | iut64 |
| 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 | iut64 |

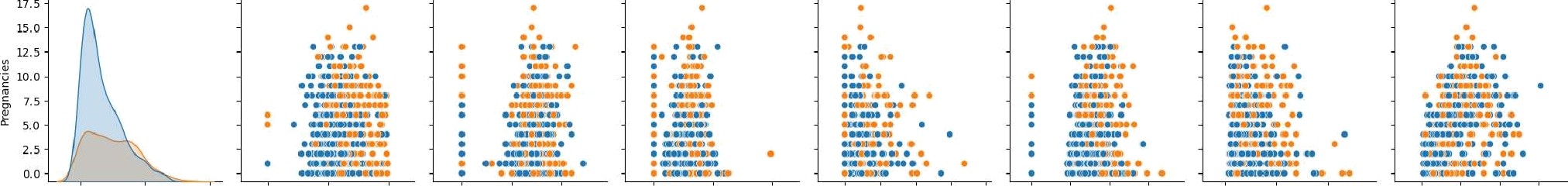
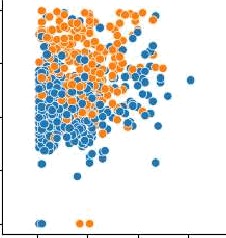
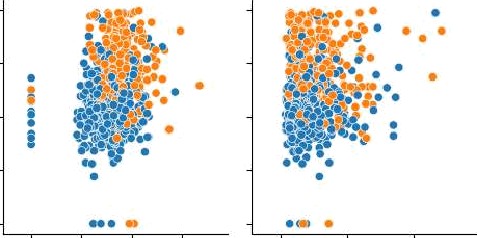
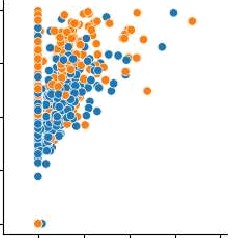
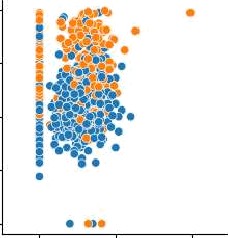
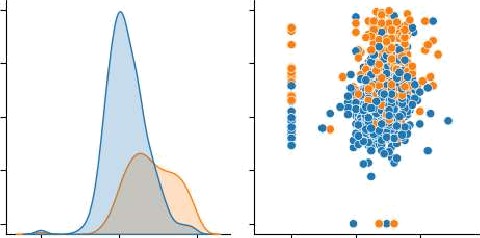
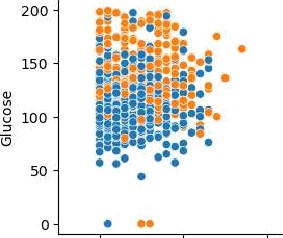
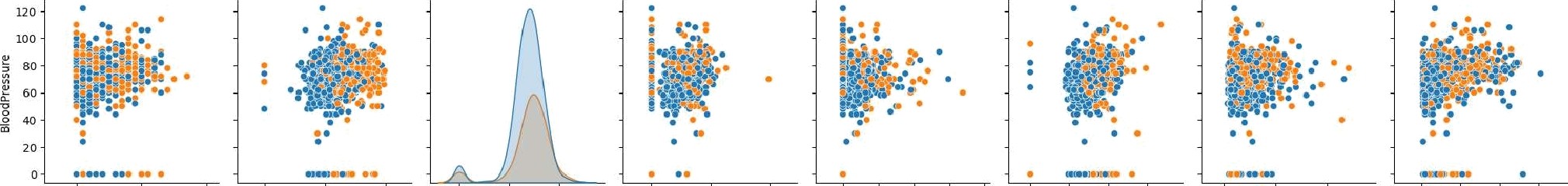
dtypes: float64(2), int64(7)

memory usage: 54.1 KB

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| dataset.describe().T | count | mean | std | m1n | 25% | 5e% | 75% | max |
| Pregnancies | 768.0 | 3.845052 | 3.369578 | 0.000 | 1.00000 | 3.0000 | 6.00000 |  |
| Glucose | 768.0 | 120.894531 | 31.972618 | 0.000 | 99.00000 | 117.0000 | 140.25000 | 199.00 |
| **BloodPressure** | 768.0 | 69.105469 | 19.355807 | 0.000 | 62.00000 | 72.0000 | 80.00000 |  |
| **SkinThickness** | 768.0 | 20.536458 | 15.952218 | 0.000 | 0.00000 | 23.0000 | 32.00000 | 99.00 |
| **Insulin** | 768.0 | 79.799479 | 115.244002 | 0.000 | 0.00000 | 30.5000 | 127.25000 |  |
| BMI | 768.0 | 31.992578 | 7.884160 | 0.000 | 27.30000 | 32.0000 | 36.60000 | 67.10 |
| DiabetesPedigreeFunction | 768.0 | 0.471876 | 0.331329 | 0.078 | 0.24375 | 0.3725 | 0.62625 |  |
| Age | 768.0 | 33.240885 | 11.760232 | 21.000 | 24.00000 | 29.0000 | 41.00000 | 81.00 |
| **Outcome** | 768.0 | 0.348958 | 0.476951 | 0.000 | 0.00000 | 0.0000 | 1.00000 |  |
| dataset . lsnuII ( ) . sum( ) |  |  |  |  |  |  |  |  |
| Pregnant ies | 0 |  |  |  |  |  |  |  |
| G1ucose | 0 |  |  |  |  |  |  |  |
| BloodPressure | 0 |  |  |  |  |  |  |  |
| Sk1nTh1c kness | 0 |  |  |  |  |  |  |  |
| Insulin | 0 |  |  |  |  |  |  |  |
| BMI | 0 |  |  |  |  |  |  |  |
| DiabetesPedigreeFunction | 0 |  |  |  |  |  |  |  |
| Age | 0 |  |  |  |  |  |  |  |



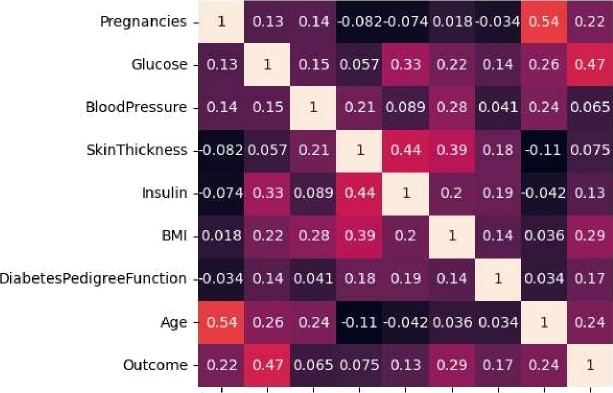
Outcome



# Heatmap

sns.heatmap(dataset.corr(), annot = True) p1t.show()

— 1.0



@ E



# Replacing zero values with Nan

SkinThic k ness

Insulin

DiabetesPedigreeF unction

dataset\_new = dataset

dataset\_new[[“Glucose“, "BloodPressure“, “SkinThickness“, “Insulin", “BMI“]] = dataset\_new[[“Glucose“, “BloodPressure“, “SkinThickness“, "Iss

# Count of NaN dataset\_new.isnull().sum()

|  |  |
| --- | --- |
| Pregnancies | 0 |
| G1ucose | 5 |
| BloodPressure | 35 |
| SkinThickness | 227 |
| Insulin | 374 |
| BMI | 11 |
| DiabetesPedigreeFunction | 0 |
| Age | 0 |
| Outcome dtype: int64 | 0 |

# Replacing NaN with mean values dataset\_new[”Glucose“].fillna(dataset\_new[”Glucose"].mean(), inplace = True)

dataset\_new[“BloodPressure“].fillna(dataset\_new[“BloodPressure“].mean(), inplace = True) dataset\_new[“SkinThickness“].fi11na(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 = Tnue)

dataset\_new.isnull().sum()

Pregnancies 0

Glucose 0

BloodPressure 0

SkinThickness 0

Insulin 0

BMI 0

DiabetesPedigreeFunction 0

Age 0

Outcome 0

dtype: int64

**CONCLUSION:**

In conclusion, our model for innovation in the AI-based Diabetes Prediction System focuses on user-centric design, multidisciplinary collaboration, data security, continuous improvement, and seamless integration with healthcare institutions. By prioritizing user needs, embracing cutting-edge technology, and upholding ethical data practices, we aim to revolutionize diabetes care globally. This model ensures accuracy, usability, and adaptability, ultimately making a positive impact on diabetes prevention and management.