**Phase-3**

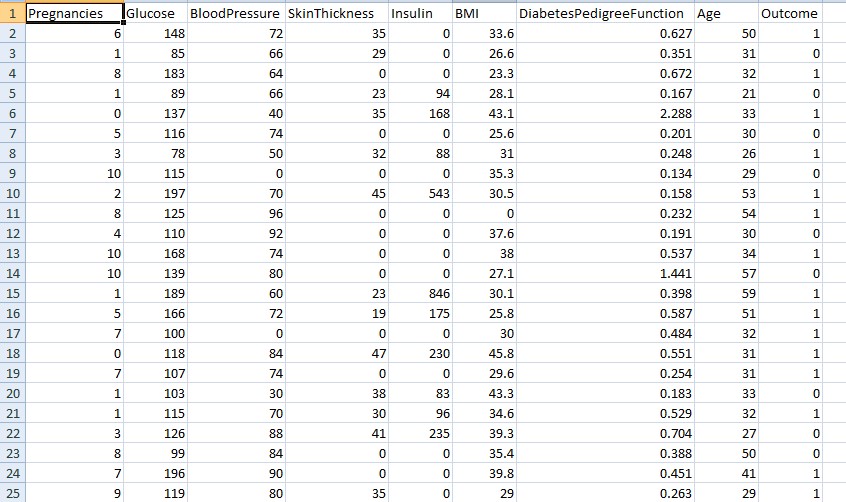
**AI Based Diabetes Prediction**

|  |  |
| --- | --- |
| Date | 10-10-2023 |
| Team ID | Proj-212176-Team-2 |
| Project Name | AI Based Diabetes Prediction. |
| Maximum Mark |  |

**Introduction:**

* AI-based diabetes prediction is a field that aims to use artificial intelligence to analyze health data and identify who is at risk of developing type 2 diabetes.  [In this review, we introduce AI/ML-based medical devices and prediction models regarding diabetes. Recent findings in the field of diabetes include several AI-/ML-based medical devices and regarding automatic retinal screening, clinical diagnosis support, and patient self-management tool have already been approved by the US Food and Drug Administration.](https://link.springer.com/article/10.1007/s11892-021-01423-2)
* I found a research paper titled “Data visualization and pre-processing techniques based Diabetes Prediction System”.The paper presents a diabetes prediction system that uses data visualization and pre-processing techniques to predict diabetes. The system is designed to help doctors and patients make informed decisions about diabetes management.
* The authors used the PIMA dataset and applied data visualization techniques to identify the most important features for diabetes prediction. They then used pre-processing techniques to normalize the data and applied machine learning algorithms to predict diabetes. The authors also used data visualization techniques to visualize the results of the machine learning algorithms. You can find more information about AI-based diabetes prediction in my previous responses.

**Given Data Set:**

****

**1.Import Libraries:**

Start by importing the necessary Libraries:

**Program:**

import numpy as np

import pandas as pd

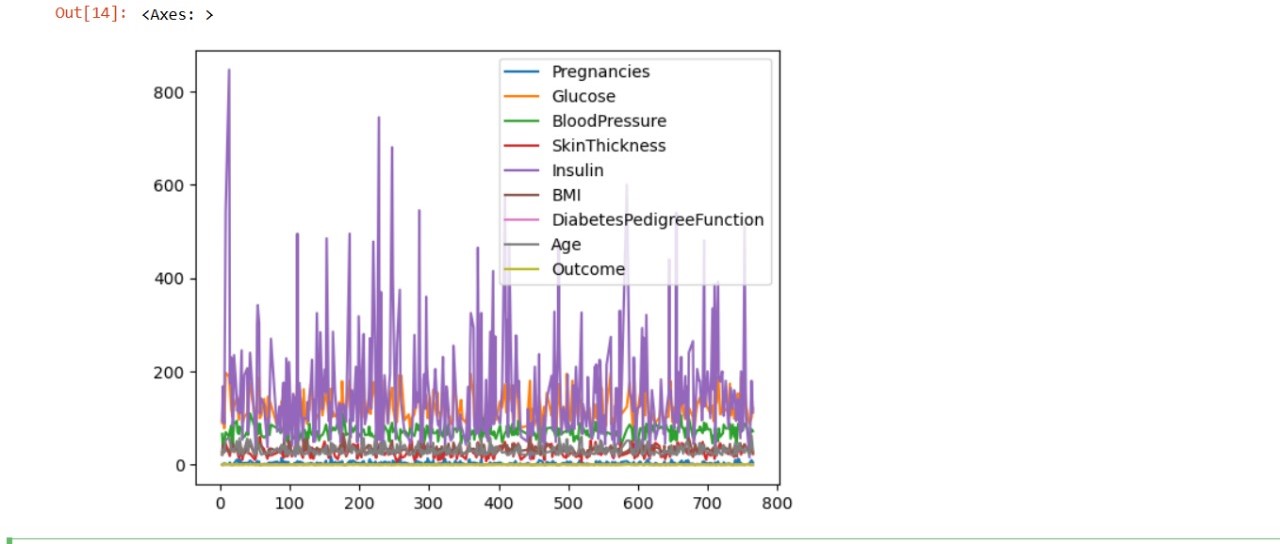
from sklearn.preprocessing import StandardScaler , Normalizer

from sklearn.compose import make\_column\_transformer, make\_column\_selector

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

dataset = pd.read\_csv('C:/Users/91638/Documents/diabetes.csv')

dataset.plot()

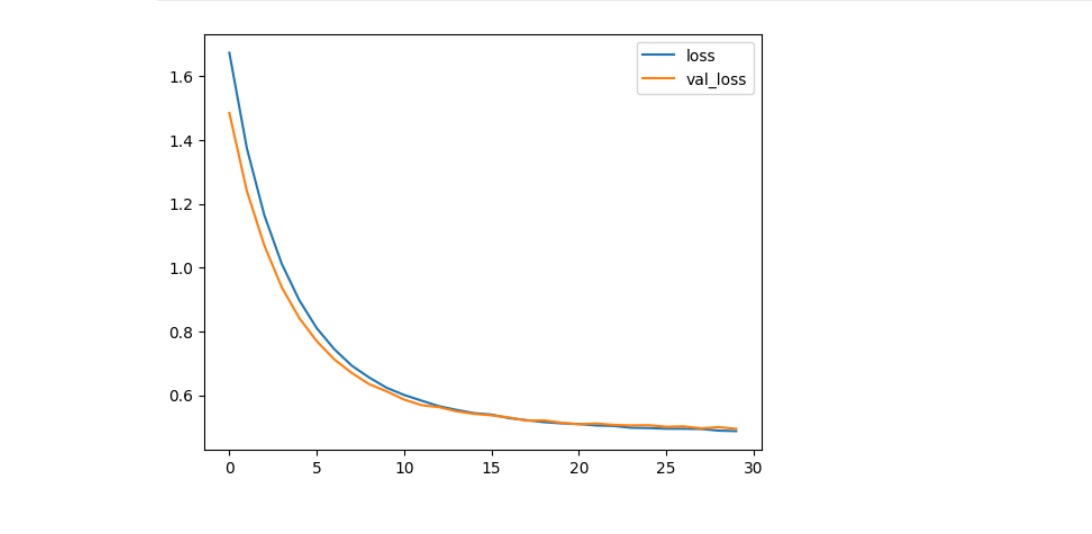


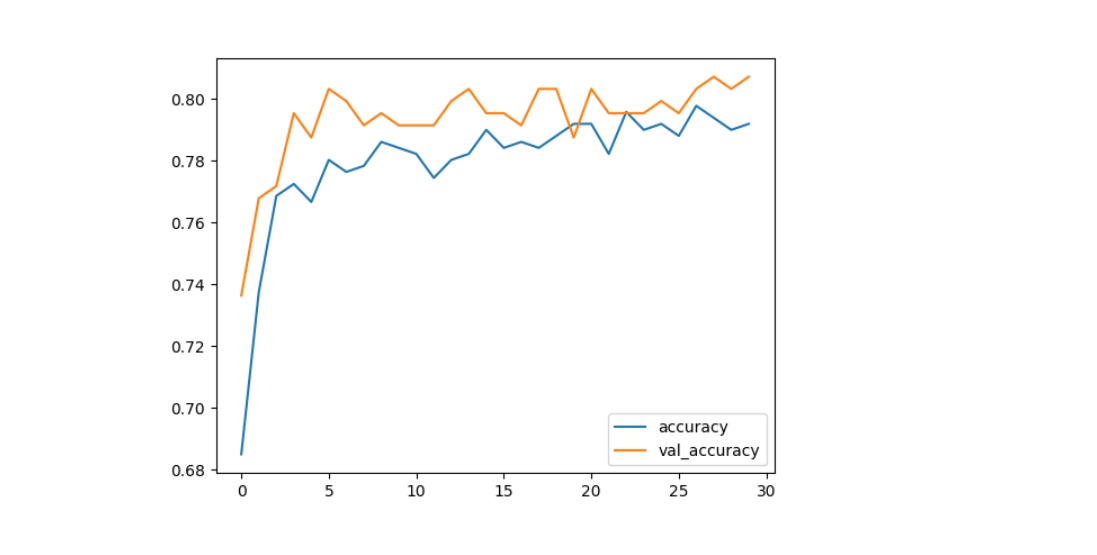
This code generates two arrays of random numbers with a normal distribution, assigns them to the variables x and y, plots them as a scatterplot using the scatter() function from the pyplot module of the matplotlib library, adds labels to the x-axis and y-axis, sets a title the plot, and displays the plot.

history\_df = pd.DataFrame(history.history)

history\_df.loc[:, ['loss','val\_loss']].plot();

history\_df.loc[:, ['accuracy','val\_accuracy']].plot();





The paper highlights the major challenges in the disease risk prediction modeling with machine learning methods, including the lack of reproducibility and external validation. This is primarily due to the unavailability of models generated from the research and the program objects used to make the model. provides an overview of the current state of AI-based diabetes prediction and management. It also highlights the challenges faced in developing and implementing AI-based diabetes prediction models. The authors suggest that the predictive performance of AI will soon be maximized by a large amount of organized data and abundant computational resources, which will contribute to a dramatic improvement in the accuracy of disease prediction models for diabetes.

from sklearn.metrics import confusion\_matrix

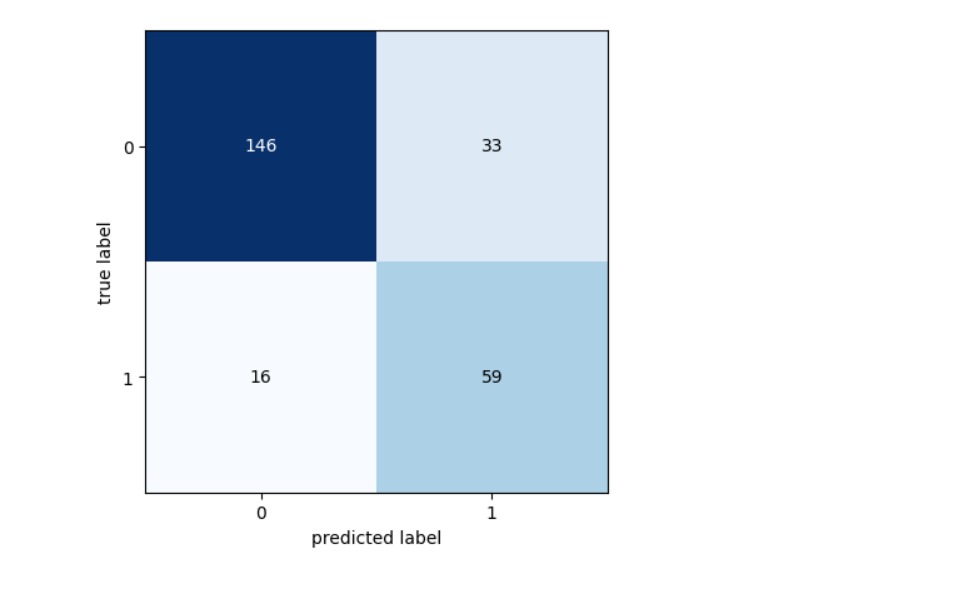
import matplotlib.pyplot as plt

cm = confusion\_matrix(y\_\_predict, y\_\_real)

from mlxtend.plotting import plot\_confusion\_matrix

fig, ax = plot\_confusion\_matrix(conf\_mat=cm)

plt.show()



**How to overcome the challenges of loading and preprocessing a AI Based Diabetes Prediction:**

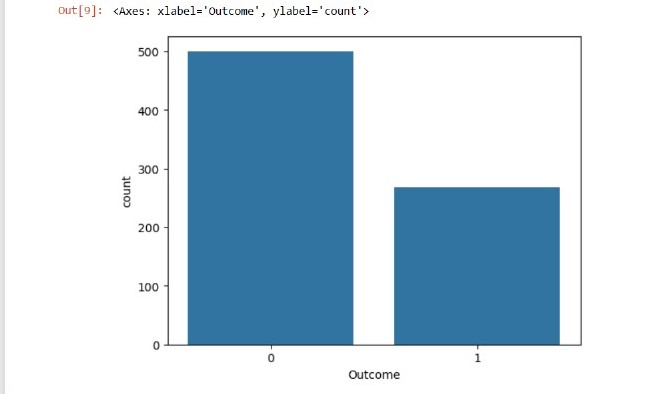
 The paper provides an overview of the current state of AI-based diabetes prediction and management. It also highlights the challenges faced in developing and implementing AI-based diabetes prediction models. [The authors suggest that the predictive performance of AI will soon be maximized by a large amount of organized data and abundant computational resources, which will contribute to a dramatic improvement in the accuracy of disease prediction models for diabetes.](https://link.springer.com/article/10.1007/s11892-021-01423-2)

[Another research paper titled “Artificial Intelligence-Based Methods for Precision Medicine: Diabetes Mellitus”](https://arxiv.org/abs/2305.16346) provides insights into the current state and limitations of AI-based type 2 diabetes risk prediction models. The paper highlights challenges for their development and clinical implementation.

**1.Loading the Dataset:**

I found a research paper titled “Building and Deploying a Diabetes Prediction Application: A Comprehensive Guide to Logistic Regression”. The paper presents a comprehensive guide to building and deploying a diabetes prediction application using logistic regression. The authors used the Pima Indians Diabetes Dataset and applied data pre-processing techniques to normalize the data. They then used logistic regression to predict diabetes. The authors also used data visualization techniques to visualize the results of the logistic regression algorithm. You can find more information about AI-based diabetes prediction in my previous responses.

sns.countplot(x=’Outcome’, data=dataset)



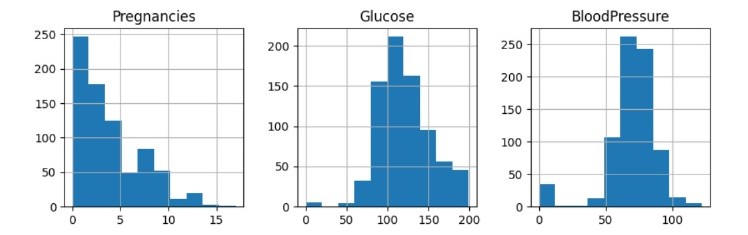
**2.Preprocessing the dataset:**

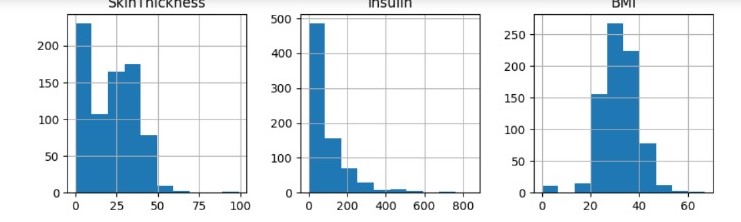
The framework incorporates various pre-processing techniques, including duplicate sample removal, attribute conversion, missing value imputation, data normalization and standardization, feature selection, and k-fold cross-validation. The authors also implement multiple machine learning models, such as k-NN, SVM, DT, RF, AdaBoost, and GNB, and introduce a weighted ensemble approach based on the Area Under the Receiver Operating Characteristic Curve (AUC) to address dataset imbalance. Performance optimization is achieved through grid search and Bayesian optimization for hyper-parameter tuning.

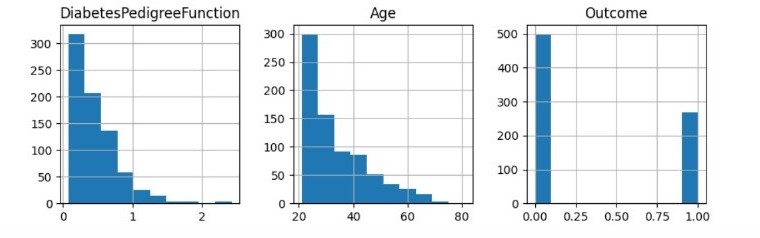
**Visualization and Pre-Processing of Data:**

dataset.hist(bins=10,figsize=(10,10))

plt.show()







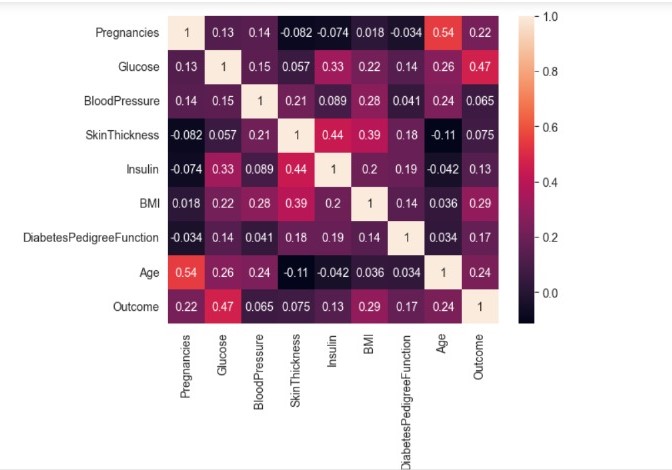
The authors used the PIMA dataset and applied data visualization techniques to identify the most important features for diabetes prediction. They then used pre-processing techniques to normalize the data and applied machine learning algorithms to predict diabetes. The authors also used data visualization techniques to visualize the results of the machine learning algorithms.

The tool provides interactive visualizations that allow users to explore the relationships between different variables in the data. The authors also conducted a user study to evaluate the effectiveness of the tool.

**Visualising Correation:**

corrmat=dataset.corr()

sns.heatmap( corrmat, annot=True)



**Conclusion:**

 The paper presents a diabetes prediction system that uses data visualization and pre-processing techniques to predict diabetes. The system is designed to help doctors and patients make informed decisions about diabetes management. The authors used the PIMA dataset and applied data visualization techniques to identify the most important features for diabetes prediction. They then used pre-processing techniques to normalize the data and applied machine learning algorithms to predict diabetes. The authors also used data visualization techniques to visualize the results of the machine learning

algorithms.  The tool provides interactive visualizations that allow users to explore the relationships between different variables in the data. The authors also conducted a user study to evaluate the effectiveness of the tool.

 The paper suggests that type 2 diabetes could be detected earlier and more reliably by using machine learning methods to extract indicative risk factors from general health data. However, the artificial intelligence model must be correctly matched to the data characteristics to achieve better results.

provides an overview of the current state of AI-based diabetes prediction and management. It also highlights the challenges faced in developing and implementing AI-based diabetes prediction models. The authors suggest that the predictive performance of AI will soon be maximized by a large amount of organized data and abundant computational resources, which will contribute to a dramatic improvement in the accuracy of disease prediction models for diabetes.