**Project Report**

**(Machine Learning)**

Implementation of **KNN, PCA, and ANN** Algorithms for Diabetes Prediction

The project aims to work with **2** machine learning algorithms :

I had work with algorithms i.e., K-Nearest Neighbours (**KNN**), Principal Component Analysis (**PCA**) , and Artificial Neural Networks(**ANN**) :

To help predict if a patient has diabetes or not.

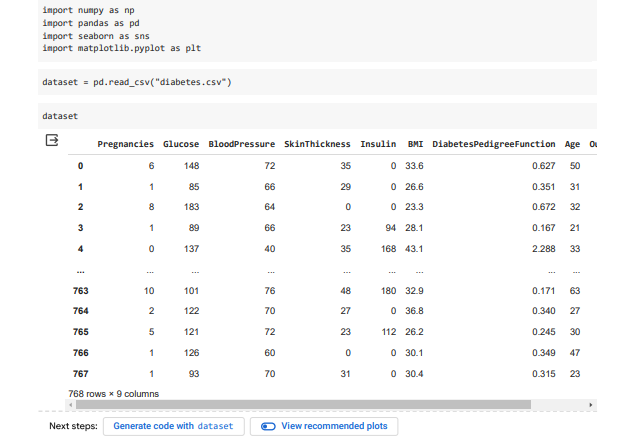
The original data is from the National Institute of Diabetes and digestive and kidney diseases , and it is used to diagnose the likelihood of a patient having the disease based on several measures.

Note that all patients are females at least 21 years old and of Pima Indian heritage.

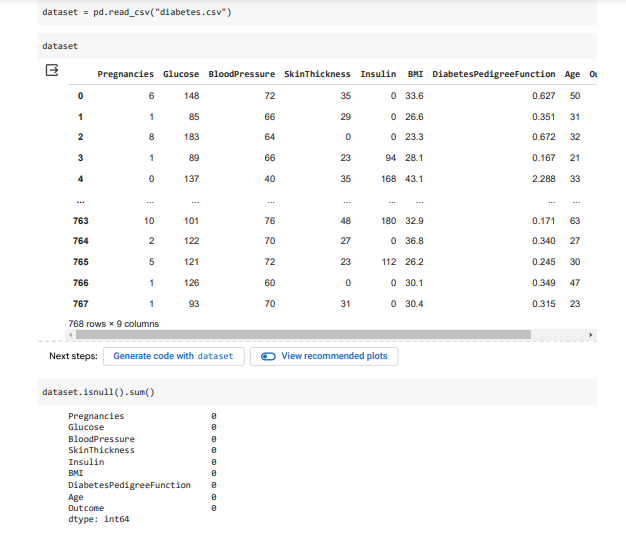
**Methodology :**

**K-Nearest Neighbors (KNN)**

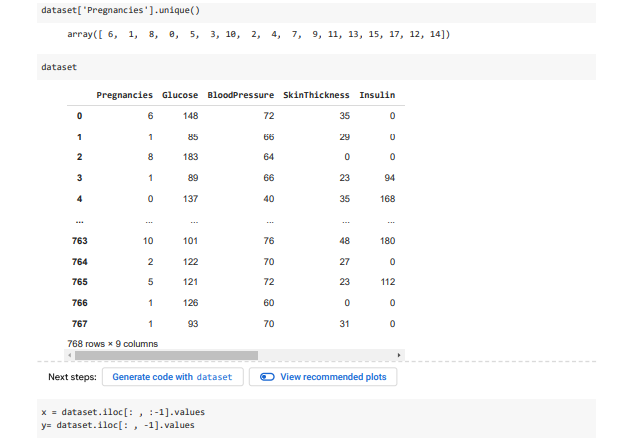
**Python Code :**



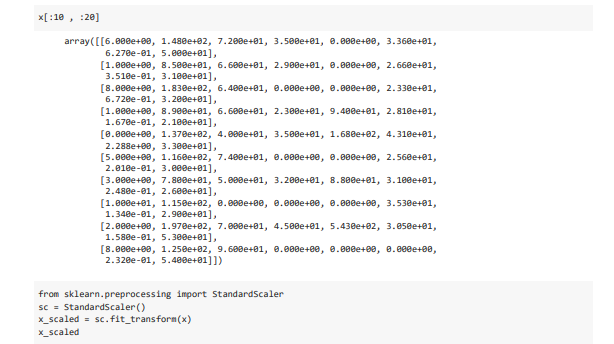
* Importing necessary libraries
* Loading dataset



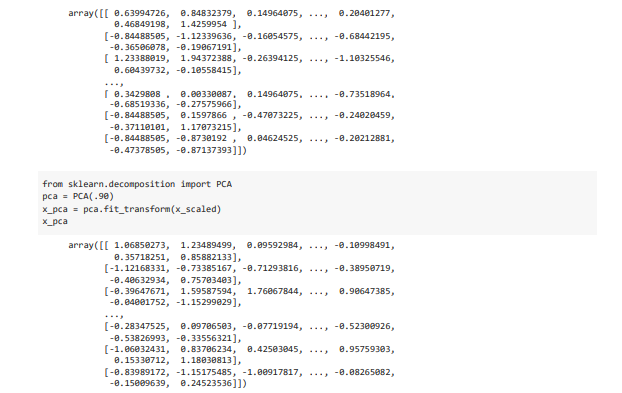
* Handling missing values
* Extracting features and labels
* Finding Unique Values of Specific Creature (Pregnancies Column).



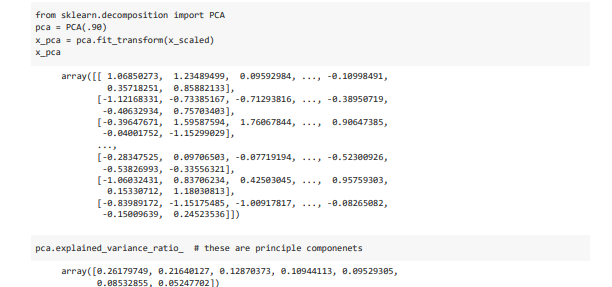
* Splitting the dataset into training and testing sets



* Feature scaling

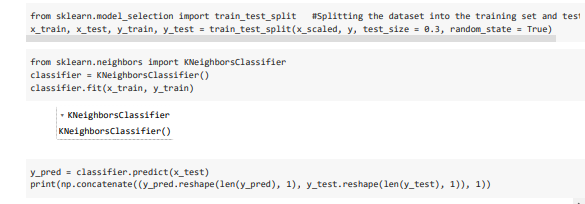


**Principal Component Analysis (PCA) :**



* Applying PCA for feature reduction
* Specify the variance to be explained(90%)
* Checking explained variance ratio

**K-Nearest Neighbors (KNN)**



* Training the KNN classifier
* Making prediction
* Evaluating performance



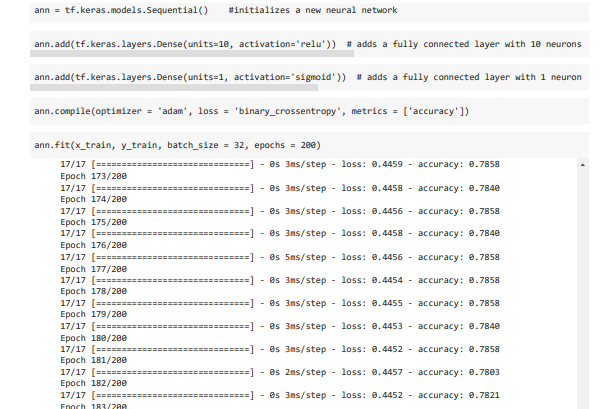
* + ***So, here 0 is our output, means that person is***

***not patient of Diabetes.***

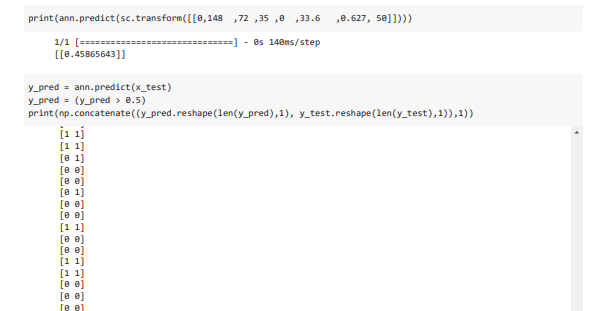
**Artificial Neural Networks (ANN)**

* Building and training the ANN





* initializes a new neural network.
* adds a fully connected layer with 10 neurons and **ReLU**(Rectified Linear Unit) helps introduce non-linearity into the model.
* adds a fully connected layer with 1 neuron and **sigmoid** activation function, typically used for binary classification tasks in neural networks.
* Making predictions.



* Evaluating performance



**Results**

The performance metrics for **KNN**, **PCA**, and **ANN** algorithms are as follows:

**K-Nearest Neighbors (KNN):**

**Confusion Matrix** : **[[128 18]**

**[ 31 54]]**

**Accuracy**: 0.7878787878787878

**Principal Component Analysis (PCA):**

**Explained Variance Ratio**: [0.26179749, 0.21640127, 0.12870373, 0.10944113, 0.09529305,0.08532855, 0.05247702]

**So, here we have 7 major Components that contributes to 90%.**

**Artificial Neural Networks (ANN):**

**Confusion Matrix: [[130 16]**

**[ 31 54]]**

**Accuracy**: 0.7965367965367965

**Conclusion**

The implementation of the **KNN**, **PCA**, and **ANN** algorithms for the prediction of diabetes based on diagnostic measurements was considered and assessed.

The results of the implementation are compared and described as follows:

* + **ANN demonstrated superior accuracy compared to KNN, achieving 79.65% against 78.79% respectively.**