**Major Project: Prediction of diabetes**

Done by: Sai Deeraj D (Verzeo ML batch 1)

This project aims to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset

**Tools used:**

Google Colaboratory - Colab is a free Jupyter notebook environment that runs entirely in the cloud. Most importantly, it does not require a setup and the notebooks that you create can be simultaneously edited by your team members - just the way you edit documents in Google Docs. Colab supports many popular machine learning libraries which can be easily loaded in your notebook.

Pandas - pandas is a software library written for the Python programming language (or) a python package for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

Scikit-learn version 0.24.2 - Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

Diabetes dataset - This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

**Algorithm:** (used k- nearest neighbor, decision tree , random forest , logistic regression , svm(support vector machine) to perform classification)

Step 1: Upload dataset

Step 2: Import pandas and read the dataset

Step 3: Convert all the data’s into 0’s and 1’s since our model works on the basis of binary classification otherwise our data won’t be understood by knn and decision tree models imported from our scikit library

Step 4: Divide our data into inputs and output

Step 5: Split our data for training and testing our model this should be done for both input and output values as well

Note: keep the testing capacity as 20 or 30 % which is standard whereas rest is used for training. While using model ensure that you have imported appropriate model from scikit library

Step 6: Now perform standardization of values if required using the respective model imported

Step 7: Import respective classifier model from scikit library

Step 8: Create an instance for the class of above model and then fit the data

Note: fitting of data must be of training values

Step 8: Now predict the output using the testing values

Step 9: Test the accuracy and get detailed analysis of your models capability`

**Conclusion:**

Prediction of diabetes is done using the above ml algorithms is done .