Predicting Diabetes using Machine Learning

Problem Statement:

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. 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.

The packages required for the project:

1.NumPy

2.Pandas

3.Seaborn

4.Matplotlib

5.SKlearn

- *IDE Used is Jupyter Notebook.*
- **Algorithm Used:**

Logistic Regression is used because according to the problem Statement Predicting whether a patient is a diabetic or not is a Binary Classification Problem.

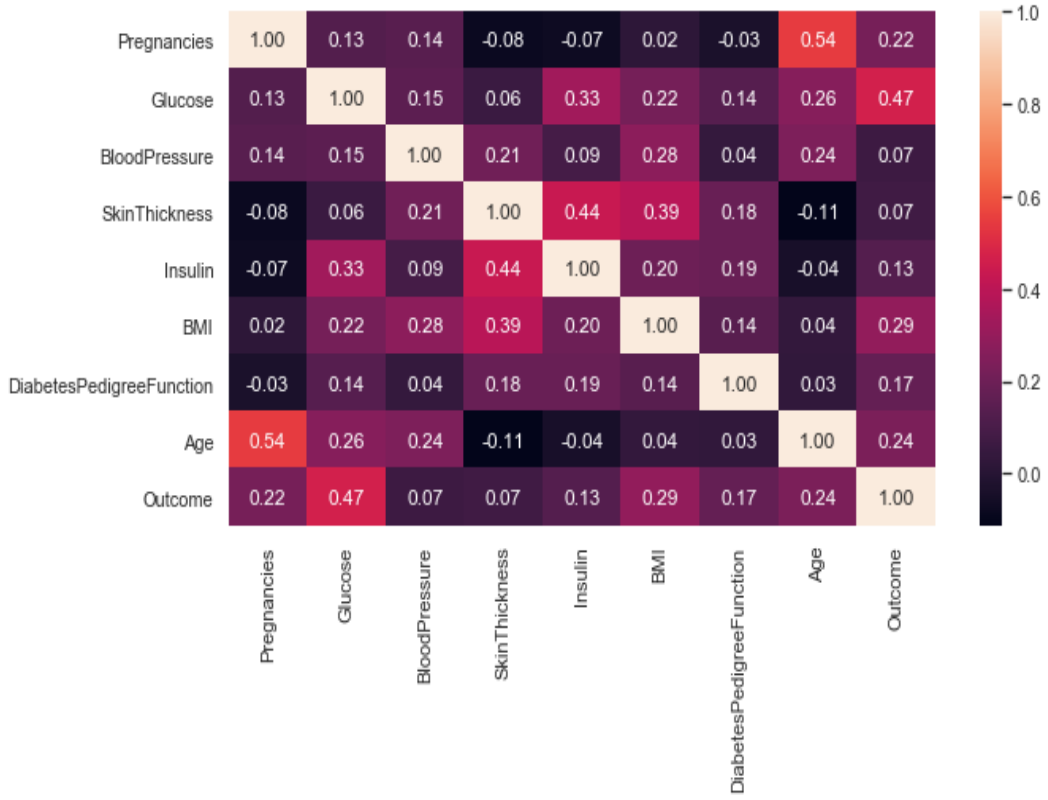
Description about the columns in the dataset

- *pregnancy:*
  - *The state of carrying a developing foetus within the female body.*
- *Glucose:*
  - *A simple sugar which is an important energy source in living organisms.*
- *Blood Pressure:*
  - *The pressure of circulating blood against the walls of blood vessels.*
- *skin Thickness:*
  - *thickness of the skin*

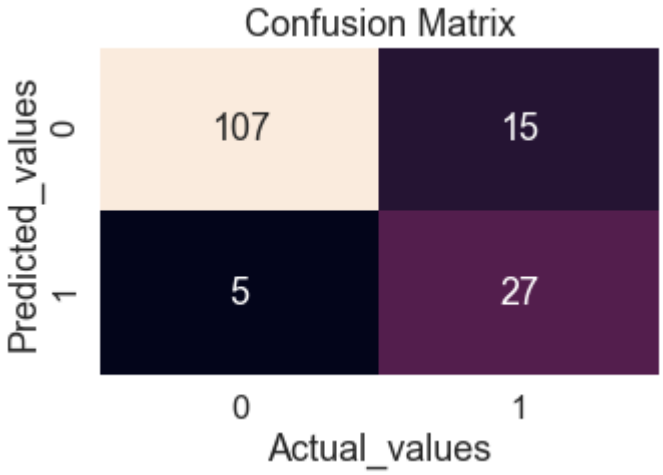
- *Insulin:*
  - *a hormone that lowers the level of glucose in the blood.*
- *BMI(BodyMassIndex):*
  - *person's weight in kilograms divided by the square of height in meters.*
- *DiabetiesPedigreeFunction:*
  - *a function which scores likelihood based on family history.*
- *Age:*
  - *The amount of time during which someone has existed.*
- *Outcome:*
  - *1 indicates the patient has diabetics whereas 0 indicates doesn't.*

➢ Code Screenshots:

- **Correlation Matrix:**



- **Confusion Matrix:**



- **Accuracy:**

jupyter miniproject Last Checkpoint: 20 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

creating an object for our model class and training the model with the training data set

```
In [32]: model=LogisticRegression()
         model.fit(x_train,y_train)
```

```
C:\Users\saibhargavi\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:432: FutureWarning: Default solver will be
changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
```

```
Out[32]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                            intercept_scaling=1, l1_ratio=None, max_iter=100,
                            multi_class='warn', n_jobs=None, penalty='l2',
                            random_state=None, solver='warn', tol=0.0001, verbose=0,
                            warm_start=False)
```

predicting the labels by using the testing data set

```
In [21]: y_preds=model.predict(x_test)
```

checking our model accuracy score (which means how well our predicted the labels).

```
In [22]: from sklearn.metrics import accuracy_score
```

```
In [23]: accuracy_score(y_preds,y_test)
```

```
Out[23]: 0.8701298701298701
```

➢ Conclusion:

*Logistic Regression Model was able to **predict the Diabetics with 87% accuracy.***