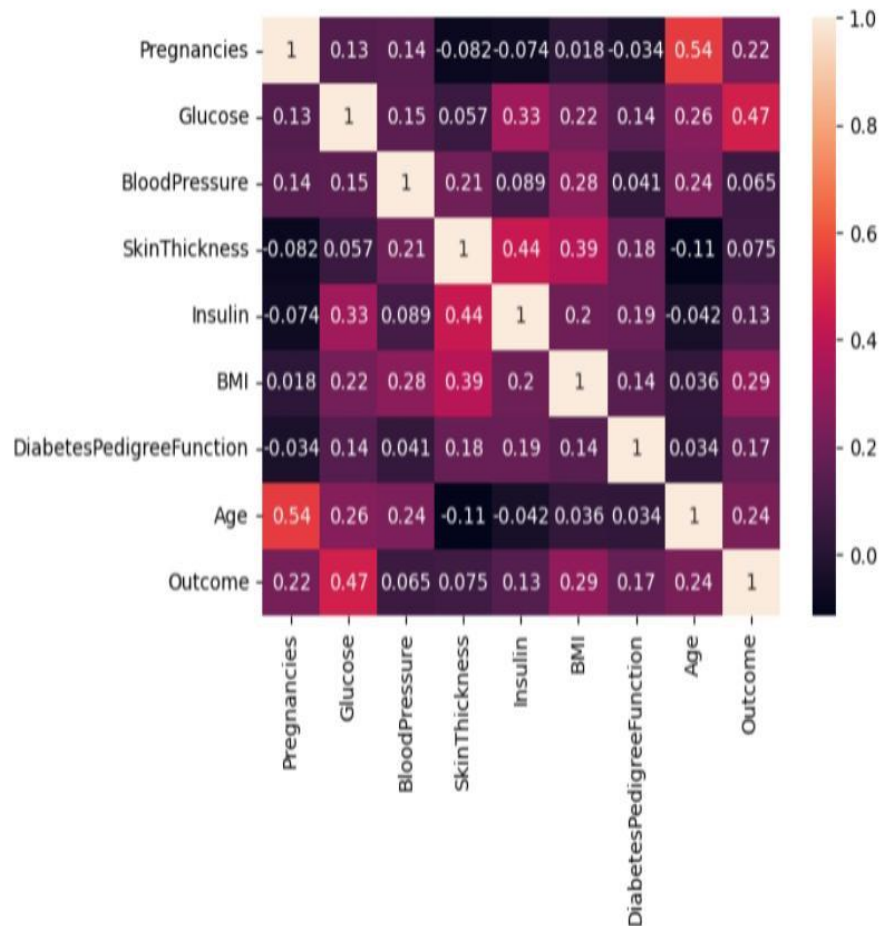


```
[ ]: sns.heatmap(dataset.corr(), annot = True)
plt.show()
```



```
[ ]: dataset_new = dataset
```

```
[ ]: dataset_new[["Glucose", "BloodPressure", "SkinThickness", "Insulin", "BMI"]] =
↳ dataset_new[["Glucose", "BloodPressure", "SkinThickness", "Insulin", "BMI"]].
↳ replace(0, np.NaN)
```

```
[ ]: dataset_new.isnull().sum()
```

```
[ ]: Pregnancies      0
      Glucose         5
      BloodPressure    35
      SkinThickness    227
      Insulin         374
      BMI             11
```

```
DiabetesPedigreeFunction    0
Age                        0
Outcome                    0
dtype: int64
```

```
[ ]: dataset_new["Glucose"].fillna(dataset_new["Glucose"].mean(), inplace = True)
dataset_new["BloodPressure"].fillna(dataset_new["BloodPressure"].mean(),
    ↳inplace = True)
dataset_new["SkinThickness"].fillna(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 = True)
```

```
[ ]: dataset_new.describe().T
```

```
[ ]:
count      mean      std      min      25%  \
Pregnancies    768.0    3.845052    3.369578    0.000    1.00000
Glucose        768.0   121.686763   30.435949   44.000   99.75000
BloodPressure  768.0    72.405184   12.096346   24.000   64.00000
SkinThickness  768.0    29.153420    8.790942    7.000   25.00000
Insulin        768.0   155.548223   85.021108   14.000  121.50000
BMI            768.0    32.457464    6.875151   18.200   27.50000
DiabetesPedigreeFunction  768.0    0.471876    0.331329    0.078    0.24375
Age            768.0    33.240885   11.760232   21.000   24.00000
Outcome        768.0    0.348958    0.476951    0.000    0.00000
```

```

50%      75%      max
Pregnancies    3.000000    6.000000   17.00
Glucose        117.000000  140.250000  199.00
BloodPressure   72.202592   80.000000  122.00
SkinThickness   29.153420   32.000000   99.00
Insulin        155.548223  155.548223  846.00
BMI             32.400000   36.600000   67.10
DiabetesPedigreeFunction  0.372500   0.626250    2.42
Age            29.000000   41.000000   81.00
Outcome         0.000000   1.000000    1.00
```

```
[ ]: from sklearn.preprocessing import MinMaxScaler
sc = MinMaxScaler(feature_range = (0, 1))
dataset_scaled = sc.fit_transform(dataset_new)
```

```
[ ]: dataset_scaled = pd.DataFrame(dataset_scaled)
```

```
[ ]: X = dataset_scaled.iloc[:, [1, 4, 5, 7]].values
Y = dataset_scaled.iloc[:, 8].values
```

```
[ ]: from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.20,
random_state = 42, stratify = dataset_new['Outcome'] )
```

```
[ ]: print("X_train shape:", X_train.shape)
print("X_test shape:", X_test.shape)
print("Y_train shape:", Y_train.shape)
print("Y_test shape:", Y_test.shape)
```

```
X_train shape: (614, 4)
X_test shape: (154, 4)
Y_train shape: (614,)
Y_test shape: (154,)
```

```
[ ]: from sklearn.linear_model import LogisticRegression
logreg = LogisticRegression(random_state = 42)
logreg.fit(X_train, Y_train)
```

```
[ ]: LogisticRegression(random_state=42)
```