```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score,precision_score,recall_score,f1_score,matthews_
import pickle


'diabetes.csv')


ancies","Glucose","BloodPressure","SkinThickness","Insulin","BMI","DiabetesPedigreeFunctio
```

```
   Pregnancies  Glucose  BloodPressure  ...  DiabetesPedigreeFunction  Age  Outcome
0            6      148             72  ...                     0.627   50        1
1            1       85             66  ...                     0.351   31        0
2            8      183             64  ...                     0.672   32        1
3            1       89             66  ...                     0.167   21        0
4            0      137             40  ...                     2.288   33        1

[5 rows x 9 columns]
(768, 9)
```

```
best_acc = 0
for i in range(50):
  predict = "Outcome"
  x = np.array(data.drop([predict],1))
  y = np.array(data[predict])
  x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.1)
  rf = RandomForestClassifier()
  rf.fit(x_train,y_train)
  y_pred = rf.predict(x_test)
  acc = accuracy_score(y_test,y_pred)
  if acc>best_acc:
    best_acc = acc
    best_model = rf

y_pred = best_model.predict(x_test)

fileName = "Diabetes.pickle"
pickle.dump(best_model,open(fileName,'wb'))

print("Accuracy =",str(best_acc*100)+"%")
```

```
Accuracy = 85.71428571428571%
```

✓ 11s completed at 00:08